APPLICATIONS OF ADVANCED CONTROL AND ARTIFICIAL INTELLIGENCE IN SMART GRIDS

EDITED BY: Qiuye Sun, Jianfang Xiao, Yonghao Gui, Dazhong Ma and Lei Xi
PUBLISHED IN: Frontiers in Energy Research

frontiers Research Topics

ISSN 1664-8714
ISBN 978-2-88976-183-8
DOI 10.3389/978-2-88976-183-8

About Frontiers

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers Journal Series

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to Quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.
Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area! Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: frontiersin.org/about/contact

APPLICATIONS OF ADVANCED CONTROL AND ARTIFICIAL INTELLIGENCE IN SMART GRIDS

Topic Editors:
Qiuye Sun, Northeastern University, China
Jianfang Xiao, Newcastle University, United Kingdom
Yonghao Gui, Aalborg University, Denmark
Dazhong Ma, Northeastern University, China
Lei Xi, China Three Gorges University, China

Citation: Sun, Q., Xiao, J., Gui, Y., Ma, D., Xi, L., eds. (2022). Applications of Advanced Control and Artificial Intelligence in Smart Grids. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-88976-183-8

Table of Contents

frontiers in Energy Research

ORIGINAL RESEARCH
published: 23 April 2021
doi: 10.3389/fenrg.2021.674943

Research on Optimal Operation of Electricity Heat Hydrogen System Based on Evaluation of New Energy Consumption Potential

Xinrui Liu, Xinying Zhao and Weiyang Zhong*

Department of Electrical Engineering, College of Information Science and Engineering, Northeastern University, Shenyang, China

Under the background of the "double high" power system, the electricity heat hydrogen system (EHHS) plays a significant role in the process of energy decarbonization. In order to meet the different optimization objectives of the system under different new energy consumption states, a new energy consumption potential assessment and optimized operation method based on intuitionistic fuzzy rough set theory is proposed. By using the intuitionistic fuzzy rough set theory, the continuous attribute data is divided into different levels and the results of its membership and non-membership are gotten at different levels. The membership results of real-time consumption data are matched with the rule sets, and then the system consumption state judgment result is obtained. In this article, the system consumption situation is divided into five states, and compared with the traditional division method, so the system state can be described more comprehensively. At the same time, the fuzzy set is used to deal with the ambiguity of the boundary between each state. The intuition theory is used to solve the problem of the uncertainty of the consumption state, and then the accurate judgment can be realized. In response to different consumption states, an optimal scheduling model is established in which a hydrogen heat energy system (HHES) is involved to meet different requirements, and a hybrid particle swarm optimization algorithm is used to solve the model. Adopting the IEEE-30 bus system as the network structure of EHHS in the simulation, the analysis shows that the dynamic state division method based on intuitionistic fuzzy rough set theory can better be used to judge the system state according to real-time variable factors. The system optimization based on the consumption state division has the advantages of improving the operating economy and increasing the consumption of new energy.

Keywords: electricity heat hydrogen hybrid system, intuitionistic fuzzy rough set, determination results of consumption state, optimal scheduling, new energy consumption

OPEN ACCESS

Edited by:
Lei Xi,
China Three Gorges University, China

Reviewed by:
Wei Hu,
Zhejiang University, China
Fei Teng,
Dalian Maritime University, China

****Correspondence:***
Xinrui Liu
liuxinrui@ise.neu.edu.cn

Specialty section:
This article was submitted to Smart Grids, a section of the journal Frontiers in Energy Research

Received: *02 March 2021*
Accepted: *19 March 2021*
Published: *23 April 2021*

Citation:
Liu X, Zhao X and Zhong W (2021) Research on Optimal Operation of Electricity Heat Hydrogen System Based on Evaluation of New Energy Consumption Potential. Front. Energy Res. 9:674943. doi: 10.3389/fenrg.2021.674943

INTRODUCTION

Due to the significant advantages of new energy, the development of wind power technology is the key direction of new energy technology in China. The output of wind power and other new energy is random and volatile, and large-scale consumption has always been a worldwide problem. The consumption of new energy is facing greater challenges due to the problems of resource scheduling

mechanisms and government subsidies (Shu et al., 2017). In recent years, the new energy industry has been developing rapidly. In view of the random interference problem of new energy, a variety of strategies have been proposed to solve the existing problems in the consumption (Xi et al., 2019, 2020). However, due to the limitation of consumption capacity in some areas, large abandonment has occurred in new energy's operation, which has affected the benefits, and whether new energy can be effectively utilized has become an urgent problem to be solved.

Many experts and scholars have carried out a lot of research work on the issue of new energy consumption. There have been studies to configure energy storage to solve the problem of new energy consumption, including electric energy storage, heat storage device, etc. In Diao et al. (2020), the complementary and coordinated relationship between energy and energy storage in the scenarios of consuming new energy and electric load peak shaving was analyzed. An integrated energy system optimization dispatch model was established to improve the capacity of new energy consumption. In Sun et al. (2019), a hierarchical dispatch framework was proposed, and wind power was consumed by dispatching electricity and heat in a cogeneration system with heat energy storage. In Chen X. et al. (2020), the coordination of power supply and heat sources was analyzed when the electric load is low during the heating period. On the premise of ensuring the maximum amount of wind power to accept and taking the lowest total operating cost as the scheduling optimization goal, a multi-source coordinated scheduling strategy was formulated. In Yang et al. (2020), the coordinated heating strategy between the electric boiler and the heat storage device was studied, and a low-carbon economic dispatching model for wind power consumption was established. In Zhou et al. (2020), the electric heat flexible load was used to further enhance the wind power consumption capacity of the system based on the flexible conversion of the source side. In Zhang Y. et al. (2020), a coordinated and optimal dispatching model of the integrated electric heat system was established, including the electric storage, the low-pressure cylinder removal, the heat storage, and the electric boiler. The rules of single operation and cooperative operation of the equipment in the above model are analyzed. The above research all studied the aspects of electric boilers with heat storage and electric storage equipment, but researchers seldom considered that electric storage equipment cannot meet the regulation requirements of the system. Therefore, the operating level of the system was still low. In Zhang et al. (2018), a two-layer optimal dispatching model of a power-natural gas integrated energy system was proposed, which considered the rational utilization of abandoned wind power by power-to-gas, and it was proved that power-to-gas could effectively improve the wind power consumption capacity through the analysis of the simulation. In Teng et al. (2019b), a coordinated optimization model of electricity heat hydrogen storage system was proposed which was more flexible and economic in grid regulation. Furthermore, it could be used to replace the battery energy storage system.

In addition to the research in the above literature, the previous optimizations were carried out with the goal of maximizing system consumption or minimizing operating cost, which did not consider the specific state of new energy fluctuations in different periods. In Yang et al. (2018), the conclusion showed that it was not appropriate to use the unified optimization method when dealing with the optimization of an integrated energy system. In Wu D. et al. (2019), the consumption method of wind power segmentation compensation was proposed, which divided the quality of wind power according to the balance cost and explored the value of peak shaving generators' contribution to the grid-connected wind power. The targeted solution model in different segments was established to promote the consumption of wind power effectively. In Ge et al. (2019), the power system was divided into normal state, alert state, and emergency state according to the amount of wind curtailment and its change trend. A staged optimal dispatch model of source-load-storage in different states was established.

This article uses intuition theory (Zhang et al., 2019; Zhan and Sun, 2020) to deal with uncertain information. The unclear boundaries of continuous data were overcome by combining with the ability of fuzzy sets (Li et al., 2020; Wei et al., 2020; Zhang P. et al., 2020). A state division method based on the consumption state of new energy is proposed. The reduction algorithm (Kumar and Prasad, 2020) is used to deal with uncertain information and massive data. Meanwhile, considering the actual situation of each consumption state and the characteristics of each optimizing adjustment equipment, a hydrogen heat energy system (HHES) including electric-hydrogen conversion, hydrogen storage, heat storage, microturbine cogeneration, and electric boiler is used in the EHHS under different consumption states to optimize and adjust.

The main innovations of this article are as follows: (1) The dynamic division method of new energy consumption state is established by using intuitionistic fuzzy rough set theory; (2) Considering the shortcomings of traditional energy storage devices and the advantages of clean hydrogen energy (Sun, 2021),the electric-to-hydrogen equipment is used in the EHHS, which combined with microturbine cogeneration and electric boiler to form the HHES to adjust the system; (3) Based on the division of the system consumption state, specific optimization goals are set to meet the actual needs in different states.

ELECTRICITY HEAT HYDROGEN SYSTEM

In this article, an EHHS is constructed, which is composed of thermal power units, wind power units, cogeneration units, and the HHES. **Figure 1** shows the overall structure of the EHHS.

The HHES is composed of five parts, which are electrolytic hydrogen unit (EHU), microturbine cogeneration unit (MCU), hydrogen storage unit (HSU), electric boiler unit (EBU) and thermal storage unit (TSU) (Teng et al., 2019a).

The energy conversion and storage relations in the HHES are as follows:

$$P_{dj,H} = \eta_{dj} \cdot P_{Ed,in} \tag{1}$$

$$P_{dg,T} = \eta_{dg} \cdot P_{Eg,in} \tag{2}$$

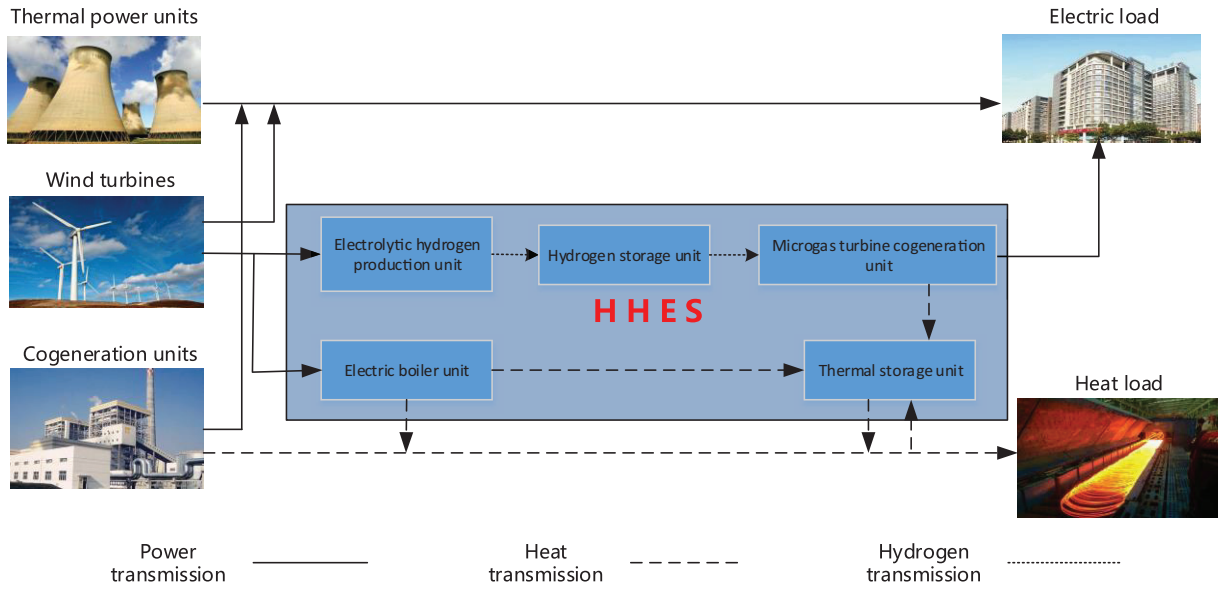

FIGURE 1 | Electricity heat hydrogen system (EHHS) structure diagram.

Where, $P_{dj,H}$ is the hydrogen production power of the EHU. $P_{dg,T}$ is the heating power of EBU. η_{dj},η_{dg} are the energy conversion efficiency of electricity to gas and electricity to heat. $P_{Ed,in}$, $P_{Eg,in}$ are the input electric power of the EHU and the input electric power of the EBU.

$$S_H(t) = [W_H(t-1) + P_{sr,H}(t) \cdot \Delta t \cdot \eta_{sr,H} - P_{sc,H}(t) \cdot \Delta t/\eta_{sc,H}]/W_{H,\max} \quad (3)$$

$$S_T(t) = [W_T(t-1) + P_{sr,T}(t) \cdot \Delta t \cdot \eta_{sr,T} - P_{sc,T}(t) \cdot \Delta t/\eta_{sc,T}]/W_{T,\max} \quad (4)$$

Where, $S_H(t)$, $S_T(t)$ are the hydrogen storage and the heat storage states. $W_H(t-1)$, $W_T(t-1)$ are the hydrogen storage and heat storage of the HSU and TSU in *t-1* period. $P_{sr,H}(t)$, $P_{sr,T}(t)$ are the input energy for HSU and TSU. $\eta_{sr,H}$, $\eta_{sr,T}$ are the energy input efficiency of the hydrogen and heat storage equipment. $P_{sc,H}(t)$, $P_{sc,T}(t)$ are the output energy of HSU and TSU. $\eta_{sc,H}$, $\eta_{sc,T}$ are the output efficiency of HSU and TSU. $W_{H,\max}$, $W_{T,\max}$ are the maximum storage capacity of HSU and TSU.

$$P_{wr,E} = \eta_{wr,E} \cdot P_{sc,H}(t) \quad (5)$$

$$P_{wr,T} = \eta_{wr,T} \cdot P_{sc,H}(t) \quad (6)$$

Where, $P_{wr,E}$, $P_{wr,T}$ are the electrical and thermal power output of the MCU. $\eta_{wr,E}$, $\eta_{wr,T}$ are the hydrogen power transfer and heat transfer efficiency of the MCU.

POWER SYSTEM OPERATION STATE DIVISION BASED ON THE LEVEL OF NEW ENERGY CONSUMPTION

Basis and Principles of System Operating State Division

Imitating the power system will have different operating states under different operating conditions. In previous studies, the power system was divided into three states which are normal state, alert state, and emergency state according to the amount of abandoned wind power (Ge et al., 2019), but in fact, it is inaccurate to delineate the consumption state by the amount of abandoned wind power, and the method dividing into three states cannot fully show the consumption state of the EHHS. Since the fuzzy boundary between each state needs to be considered when dividing, the problem of control failure caused by the inaccuracy of the original division requires the use of intuitionistic fuzzy rough set (Pawlak, 1982; Chen Y. et al., 2020) to solve. Combining the adjustment capabilities of each unit in the HHES and the amount of unconsumed wind power, the EHHS is divided into five states: very suitable (VS), moderately suitable (MS), generally suitable or unsuitable (GSU), moderately unsuitable (MU) and very unsuitable for consumption (VU). The division result is shown in **Figure 2**.

The Method of Consumption States Division Based on Intuitionistic Fuzzy Rough Set Theory

Attribute Data of State Division

Decision attributes that influence the division of state

Decision attribute is the key factor that determines the state of consumption. Based on the structure of the EHHS, the

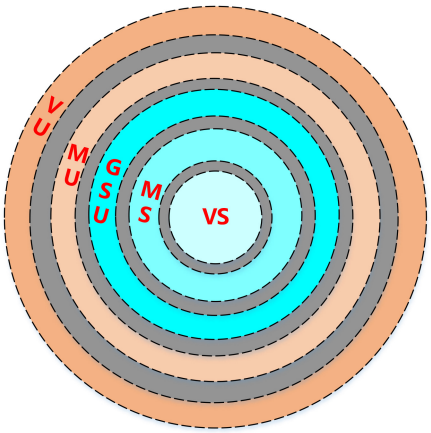

FIGURE 2 | Schematic diagram of state division.

TABLE 1 | Conditional attributes of state division.

Category	Conditional attributes
Meteorological data	Room temperature, outdoor temperature, wind speed, wind direction, etc
Time data	Time of use, month, etc
Social data	Electricity price

proportion of wind power consumption by HHES (η) and proportion of unit output (δ) are selected as the decision attributes.

$$\eta = \frac{P_p^t - P_r^t}{P_{f,\max}} \times 100\% \tag{7}$$

Where, P_p is the predicted wind power output. P_r is the actual wind power consumed. $P_{f,\max}$ is the maximum historical wind power output.

The higher the η, the more wind power the HHES needs to consume, and the more the system tends to be unsuitable for consuming wind power. When η is close to 0, the system consumption state cannot be further delineated by η. The system state needs to be delineated by δ.

$$\delta = \frac{P_h^t + P_{rd}^t}{P_d^t} \tag{8}$$

Where, P_d is the electrical load of the system.

Conditional attributes that relate to state division

This article selects a series of data such as weather and load as the conditional attributes. The result classified by data categories is shown in **Table 1**.

Intuitive Fuzzification of Attribute Data

Intuitionistic fuzzification of conditional attributes

There are two types of attribute data: discrete data and continuous data. For discrete data, rough set can process it directly, while for continuous data, it must be transformed. Taking the wind speed attribute as an example, the results of classification of membership are shown in **Figure 3**. The attribute data is divided into five fuzzy sets including extremely low (EL), low (L), moderate (M), high (H) and extremely high (EH). The maximum and minimum of historical wind speed are set as a_i^+, a_i^-, and the remaining are determined according to the statistical law that μ= 1 accounted for 20% in each rough set.

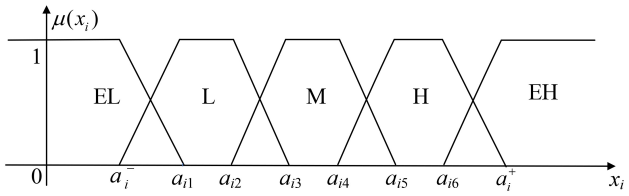

FIGURE 3 | Trapezoidal membership function.

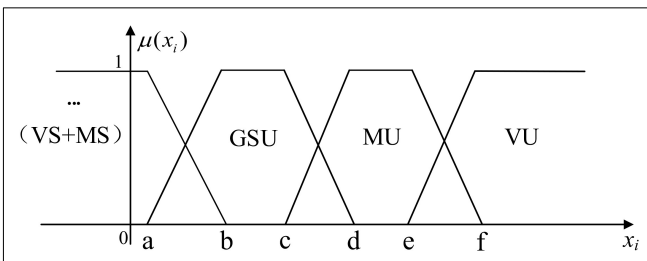

FIGURE 4 | Trapezoid membership function of decision attribute η.

Taking moderate wind speed as an example, its trapezoidal membership function is:

$$\mu_{Moderate\ wind\ speed}(x_i) = \begin{cases} \frac{x_i - a_{i2}}{a_{i3} - a_{i2}}, & a_{i2} \le x_i \le a_{i3} \\ \frac{a_{i5} - x_i}{a_{i5} - a_{i4}}, & a_{i4} \le x_i \le a_{i5} \\ 1, & a_{i3} \le x_i \le a_{i4} \\ 0, & other \end{cases} \tag{9}$$

For conditional attribute data, which has sufficient historical data records, the intuitive index is small.

The statistical rules of historical data can be given by experts.

Intuitionistic fuzzification of decision attributes

The membership function image of the decision attribute η is shown in **Figure 4**. $d = \frac{P_{dg.\max}}{P_{f,\max}}, f = \frac{P_{dg.\max} + P_{dj,\max}}{P_{f,\max}}$. The statistical law that μ= 1 accounted for 20% in each rough set, and the boundaries of fuzzy sets can be obtained according to the above law. $P_{dg,\max}$ is set to be approximately $2P_{dj,\max}$. $P_{dg,\max}$ is the maximum wind power that can be consumed by the EBU. $P_{dj,\max}$ is the maximum wind power that can be consumed by the EHU.

Through the decision attributes η, we can get the membership situation in GSU, MU, VU. When the system is in the consumption state of VS and MS, all wind power in the system can be consumed. So, it is impossible to determine its specific consumption state. The decision attribute δ was used to further determine the consumption state through P_h and P_{rd} .

The decision attribute δ is used to determine the consumption state, as shown in **Figure 5**. $x = \frac{P_{h,\min} + P_{rd,\min}}{P_d^t}, k = \frac{P_{h,\max} + P_{rd,\max}}{P_d^t}$.

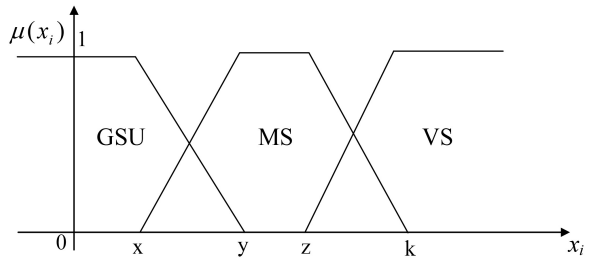

FIGURE 5 | Trapezoidal membership function of decision attribute δ.

The demarcation is related to P_d, and the situation boundary is determined according to the statistical law that μ= 1 accounted for 20% in each rough set.

Because the boundaries of each fuzzy set of decision attributes are related to the maximum and minimum output of equipment, the intuitive index is small, which is generally less than 0.1.

Construction of State Division Rule Set

Attribute reduction of decision table

Condition attributes and decision attributes can be selected to form a decision table, then the decision table needs to reduce the attributes. The dependence is used to determine the relationship between each condition attribute and decision attribute. The decision table after reduction is obtained by removing conditional attributes which are less dependent and keeping conditional attributes which are more dependent on decision attributes (Ren and Xue, 2018; Wu Y. et al., 2019).

Value reduction of decision rules

After attributes reduction, there are still a lot of redundancy in the decision table, so it is necessary to remove the redundancy by reducing the value of the rule set.

The Determination of the Consumption State

After the rule set is constructed, the real-time data that have undergone intuition fuzzification are matched with the obtained historical evaluation rules to determine the real-time consumption state. In the matching process, if the matching degree is less than the setting threshold value, then the real-time data information matches the evaluation rule completely, and membership degree results of decision attributes can be obtained. Otherwise, it will return to rematch. The real-time assessment data are updated to the historical database. In the process of judging the membership of decision attributes, if $\eta \geq b$, the state matching results can be directly obtained. The membership result of η is the elementary result of the situation determination. If not, the consumption state matching results need to be obtained through δ. As the HHES is involved in consumption, it is necessary to consider the consumption potential of the energy storage units. Combing the intuitive coefficient, the consumption state will be determined.

For example, when the system consumption state is GSU, it needs to be combined with real-time HSU data to determine whether its remaining capacity can meet the current consumption demand. If it is satisfied, the consumption state is obtained. If not, it needs to be determined whether it meets the next state.

OPTIMAL SCHEDULING OPERATION BASED ON CONSUMPTION STATE DIVISION METHOD

Optimal Dispatch Model of EHHS Based on Division Method

The system optimization objectives and constraint requirements for each consumption state are shown in **Figure 6**.

Optimization Model of VS System

Optimization objective function

In this state, the system can absorb all the wind power. The HSU and MCU can be used together to reduce the unit output. Economic operations should be implemented to reduce operating costs.

The operation costs include the generating cost of cogeneration units, the generating cost of thermal power units, and the work scheduling cost of the HSU and MCU. The generation cost of cogeneration and thermal power units include the operation cost and start cost (Liu et al., 2015).

$$\min\ C_t = C_c + C_h + C_q + C_{\mathrm{wr}} \tag{10}$$

Where, C_c, C_h, C_q, C_{wr} are the generation cost of cogeneration units, thermal power units, the HSU, and the MCU, respectively.

$$C_c = \sum_{t\in T_n}\sum_{i=1}^{I_c}[a_i(P_{c,i}^t)^2 + b_iP_{c,i}^t + c_iP_{c,i}^tH_{c,i}^t + d_i(H_{c,i}^t)^2 + e_iH_{c,i}^t + f_i] + [\sum_{t\in T_n}\sum_{i=1}^{I_c} v_{i,t}(1-v_{i,t-1})Q_i] \tag{11}$$

Where, T_n represents the number of scheduling periods in VS. I_c is the number of all cogeneration units. a_i, b_i, c_i, d_i, e_i, f_i are the coal consumption coefficients of the cogeneration unit. $P_{c,i}^t$ is the electric output of the cogeneration unit i. $H_{c,i}^t$ is the heat output of the cogeneration unit i. $v_{i,t}$ is the operating state of unit i at time t, $v_{i,t} = 1$ represents the operation of the unit, $v_{i,t} = 0$ represents the shutdown of the unit. Q_i is the start cost of a conventional unit i.

$$C_h = \sum_{t\in T_n}\sum_{i=1}^{I_h}\{(\alpha_i(P_{h,i}^t)^2 + \beta_iP_{h,i}^t + \gamma_i) + [\sum_{t\in T_n}\sum_{i=1}^{I_h} u_{i,t}(1-u_{i,t-1})S_i]\} \tag{12}$$

Where, I_h is the number of all thermal power units. α_i,β_i,γ_i are the coal consumption coefficient of thermal power units. $u_{i,t}$ is the operating state of unit i in t period, $u_{i,t} = 1$ represents the operation of the unit, $u_{i,t} = 0$represents the shutdown of the unit.

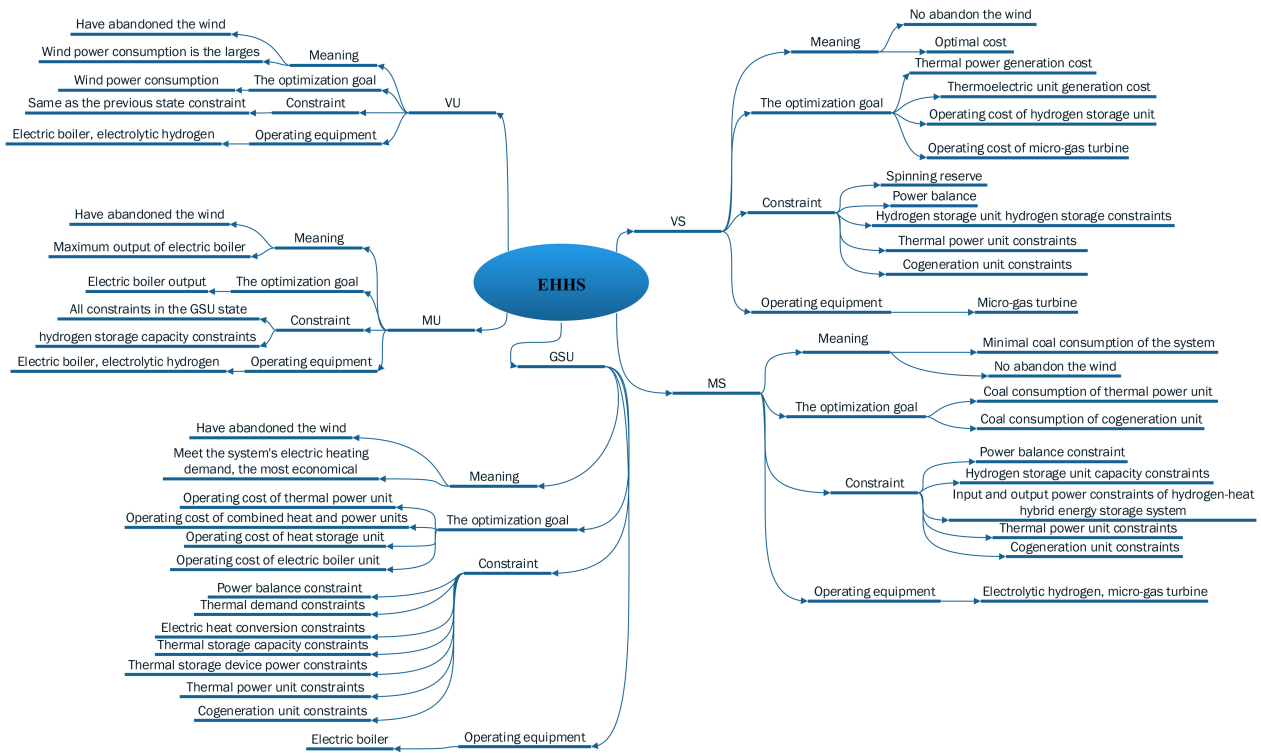

FIGURE 6 | System optimization objectives and constraints under each consumption state.

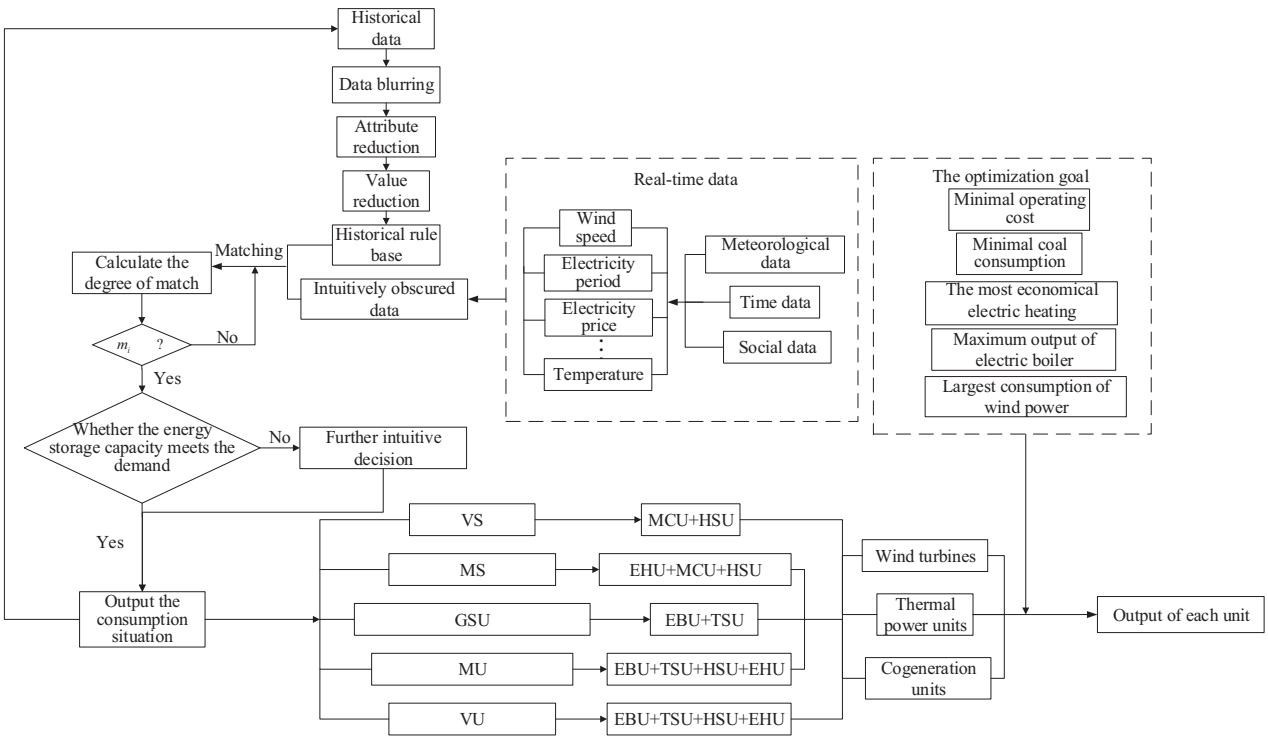

FIGURE 7 | Optimize the scheduling process.

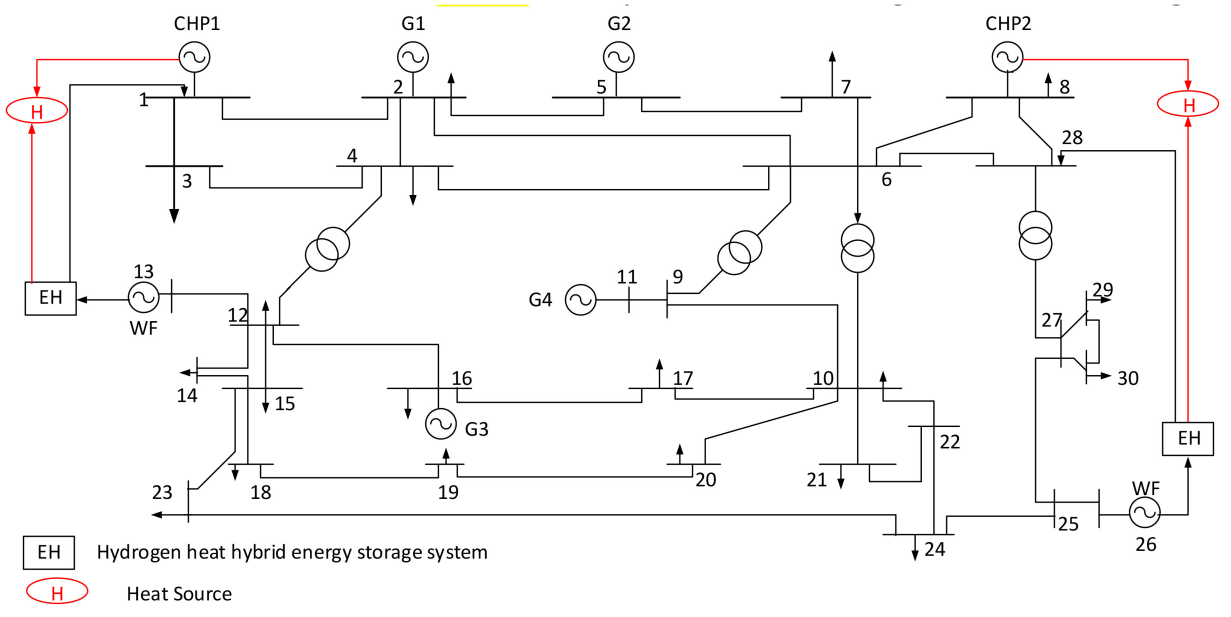

FIGURE 8 | System structure topology diagram.

S_i is the start cost of a conventional unit i.

$$C_q = \sum_{t \in T_n} C_{q,w} \cdot \left| P_q(t) \right| \tag{13}$$

Where, $C_{q,w}$ is the unit maintenance cost of the HSU. $P_q(t)$ is the power of hydrogen storage and discharging in time period t, $P_q(t) > 0$ represents energy charging, $P_q(t) < 0$ represents energy discharging.

$$C_{wr} = C_h + C_{wq} = \frac{C_{H_2}}{L_{H_2}} \sum_{t \in T_r} [\frac{P_{E,out}(t)}{\eta_{wr,E}} \Delta t] + \sum_{t \in T_r} [U_{i,t}(1 - U_{i,t-1})] C_{qd} \tag{14}$$

Where, C_h is the hydrogen cost of the MCU. C_{wq} is the start cost of the MCU. C_{H_2} is the unit price of hydrogen. L_{H_2} is the low calorific value of hydrogen, which is generally 3.1 kW.h/m^3. T_r is the number of running periods of the MCU. $P_{E,out}(t)$ is the discharge power of the HHES. $\eta_{wr,E}$ is the hydrogen-electricity conversion efficiency of the MCU. Δt is the time interval of each period. C_{qd} is the start cost of the MCU.

Constraints

Ignoring the network loss, the power balance constraints satisfy the following formula:

$$\sum_{i}^{I_h} P_{h,i}^t + P_f^t + \sum_{i}^{I_c} P_{c,i}^t + P_{E,out}^t = P_d^t \tag{15}$$

$$H_c^t + P_{T,out}^t = H_r^t \tag{16}$$

Where, H_r^t is the heat load value during t period. $P_{E,out}$ and $P_{T,out}$ are the output electricity and heating power of the HHES.

Thermal power units constraints:

The uncertainty of wind power output leads to the increase of system randomness. Positive and negative reserve capacity are used to eliminate the wind power prediction error:

$$\sum_{i=1}^{I_h} \min(P_{h,i,\max}^t - P_{h,i}^t, r_{i,u} T_3) \geq P_d^t \times L\% + P_f^t \times f_u\% \tag{17}$$

$$\sum_{i=1}^{I_h} \min(P_{h,i}^t - P_{h,i,\min}^t, r_{i,d} T_3) \geq (P_{f,\max} - P_f^t) \times f_d\% \tag{18}$$

Where, T_3 is the rotation standby response time. $L\%$ is the demand for positive rotation reserve due to the load prediction error. $f_u\%$, $f_d\%$ are demands of positive and negative rotation reserve for prediction error of wind power output. $P_{f,\max}$ is the maximum output of wind power.

$$P_{h,i,\min} \leq P_{h,i}^t \leq P_{h,i,\max} \tag{19}$$

$$r_{i,\mathrm{d}} \Delta t \leq P_{h,i}^t - P_{h,i}^{t-1} \leq r_{i,\mathrm{u}} \Delta t \tag{20}$$

TABLE 2 | Grid installed capacity.

Unit type	Installed capacity/minimum technical output (MW)	The proportion (%)
Thermal power units	2 × 10,000; 2 × 8,000 2 × 6,000; 2 × 4,000	83.3
Cogeneration units	2 × 450; 2 × 100	2.3
Wind turbines	5,500	14.3

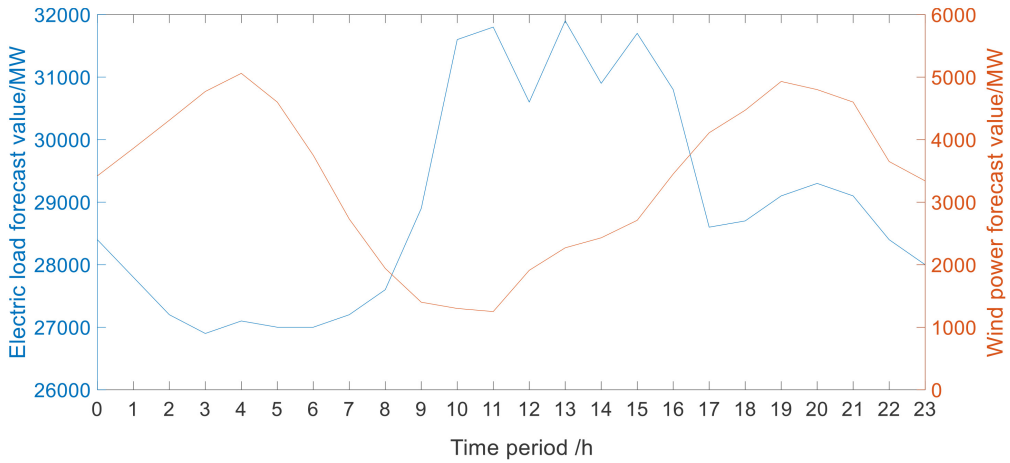

FIGURE 9 | Electric load and wind power output forecast curve.

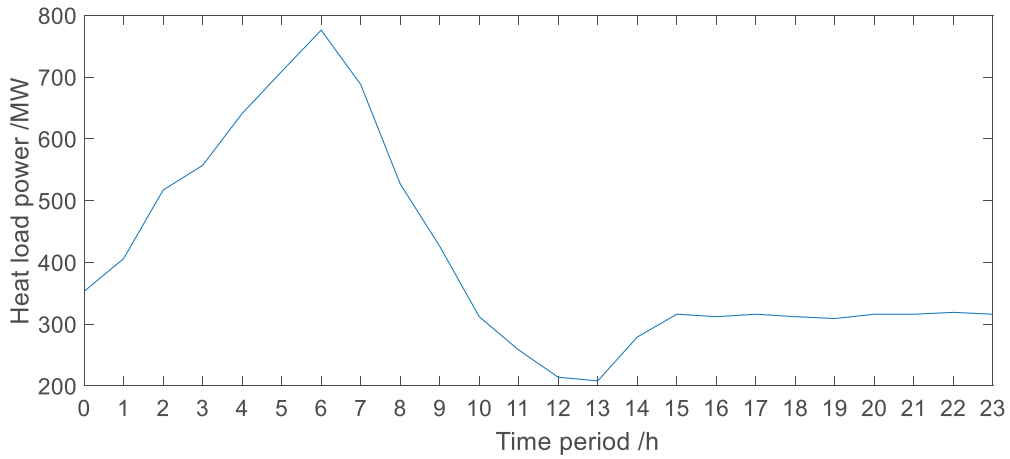

FIGURE 10 | Heat load curve.

TABLE 3 | Historical data of state division.

Data (Scheduling period)/1 h	Wind speed P1/m/s	Wind direction P2/angle	Electricity period P3/t	Electricity price P4/yuan	Outdoor temperature P_5/°C	Room temperature P6/°C	Month P7/t	$_{\eta}$ Q_1/%	$_{\delta}$ Q_2
1	1.98	20.5	09:15	0.7	−18	21.2	12	0	0.944
2	2.93	67.5	14:30	0.7	−15.5	22.4	1	0	0.941
3	3.15	87.5	16:15	0.7	−14.2	22.6	2	0	0.807
4	2.1	291.5	12:45	0.7	−10.2	23.4	1	0	0.94
5	1.38	270	18:30	0.7	−11.3	23	1	5.4	0.836
6	2.44	332	22:00	0.3	−10.5	22.8	1	0	0.865
7	3.81	240	02:15	0.3	−16.6	22.3	12	47	0.907
8	3.37	90	07:45	0.5	−13	22.7	1	0	0.924
98	4.25	110	13:15	0.7	−17.4	21	1	0	0.925
99	4.82	180	03:15	0.3	−22.8	19.6	12	52.1	0.908
100	4.73	240	19:15	0.5	−19.2	20.8	12	1.3	0.837

Cogeneration units constraints:

$$P_c^t = N_c H_c = N_c(H_r^t - P_{T,out}) \tag{21}$$

$$P_{c,\min} \le P_c^t \le P_{c,\max} \tag{22}$$

$$h_{i,\mathrm{d}}\Delta t \le P_{c,i}^t - P_{c,i}^{t-1} \le h_{i,\mathrm{u}}\Delta t \tag{23}$$

Where, P_c^t is the power output of cogeneration units. N_c is the heat-to-electricity ratio of the cogeneration unit. $P_{c,\min}$ and $P_{c,\max}$ are the minimum and maximum outputs of the i-th thermal power unit, $P_{c,\min} = 0$.

Constraint of the HHES:

$$\begin{cases} 0 \le R_{cq}^t \le R_{cq,max} \\ 0 \le \frac{P_{E,out}^t}{\eta_{wr,E}\cdot\eta_{sc,H}} \cdot \Delta t \le R_{cq}^t \end{cases} \tag{24}$$

Where R_{cq}^t is the hydrogen storage capacity of the HSU in period t. $R_{cq,max}$ is maximum storage capacities of the HSU.

Optimization Model of MS System

Optimization objective function

In this state, although the system can fully consume wind power, the output of conventional units is close to the minimum. In order to avoid the frequent climbing of units and curb the fluctuation of wind power, the EHU, HSU, and MCU work at this state. The system is optimized with the goal of minimizing the coal consumption of the EHHS.

$$\begin{aligned} \min f = \; & C_{c,m} + C_{h,m} = \sum_{t\epsilon T_v}\sum_{i=1}^{I_c}(a_i(P_{c,i}^t)^2 + b_i P_{c,i}^t + c_i P_{c,i}^t H_{c,i}^t \\ & + d_i(H_{c,i}^t)^2 + e_i H_{c,i}^t + f_i) + \sum_{t\epsilon T_v}\sum_{i=1}^{I_h}(\alpha(P_{h,i}^t)^2 \\ & + \beta_i P_{h,i}^t + \gamma_i) \end{aligned} \tag{25}$$

Where, T_v represents the number of scheduling periods in MS. $C_{c,m}$ is the coal consumption of cogeneration units. $C_{h,m}$ is the coal consumption of thermal power units.

Constraints

Ignoring the network loss, the power balance constraints satisfy the following formula:

$$\sum_i^{I_h} P_{h,i}^t + P_f^t + \sum_i^{I_c} P_{c,i}^t + P_{E,out}^t = P_d^t \tag{26}$$

$$H_c^t + P_{T,out}^t = H_r^t \tag{27}$$

$P_{E,out} > 0$, the EHU works. $P_{E,out} < 0$, the MCU works.

The HSU capacity constraints:

$$R_{cq,\min}^t \le R_{cq}^t \le R_{cq,\max}^t \tag{28}$$

$$R_{cq}^t = R_{cq}^{t-1} + P_{dj,H}^t \cdot \Delta t \cdot \eta_{sr,H} - P_{E,out}^t \cdot \Delta t/\eta_{sc,H} \tag{29}$$

Where, R_{cq}^t is the hydrogen storage capacity of the HSU in period t. $R_{cq,\min}^t$ and $R_{cq,\max}^t$ are minimum and maximum storage capacities of the HSU in period t.

Input and output power constraints of the HHES:

$$0 \le P_{E,in}^t \cdot \eta_{dj} \cdot \eta_{sr,H} \cdot \Delta t \le R_{cq,\max} - R_{cq}^{t-1} \tag{30}$$

$$0 \le P_{E,out}^t \cdot \Delta t \le R_{cq}^t \cdot \eta_{sr,H} \cdot \eta_{wr,E} \tag{31}$$

Where, $R_{cq,\max}$ is the maximum hydrogen storage capacity of the HSU.

Constraints of thermal power units and cogeneration units are same as the constraints in VS.

Optimization Model of GSU System

Optimization objective function

In this state, the system cannot consume all the wind power. The EBU starts to work, which has to consume the wind power to satisfy the heat demand. The optimizing target is the highest economy that meets the electric and heat demand.

$$\min J = C_h + C_c + C_{cr} + C_{dg} \tag{32}$$

Where, C_{cr} is the operating cost of the TSU. C_{dg} is the operation cost of the EBU.

$$C_{cr} = \sum_{t\in T_m} C_{cr,w} \cdot |P_r(t)| \tag{33}$$

$$C_{dg} = C_{dh} + C_q = \sum_{t\in T_m}(\frac{P_{dg,T}^t}{\eta_{dg}} \cdot S_{dg}) + \sum_{t\in T_m}[G_{i,t}(1 - G_{i,t-1})]C_{gq} \tag{34}$$

Where, T_m represents the number of scheduling periods in GSU. $C_{cr,w}$ is the maintenance cost of the TSU. $P_r(t)$ is the heat storage and release power in time period t, $P_r(t) > 0$ represents energy charging, $P_q(t) < 0$ represents energy release. C_{dh} is the electric to heat cost of the EBU. C_q is the start cost of the EBU. S_{dg} is the EBU conversion unit cost. $G_{i,t}$ is the EBU operating state, 1 represents operation, 0 represents shutdown. C_{gq} is the starting cost of the EBU once.

Constraints

Ignoring the network loss, the power balance constraints satisfy the following formula:

$$\sum_i^{I_h} P_{h,i}^t + P_f^t + \sum_i^{I_c} P_{c,i}^t = P_d^t + P_{E,in}^t \tag{35}$$

$$H_c^t + P_{T,out}^t = H_r^t \tag{36}$$

Where, $P_{E,in}^t$ is the power output of the HHES.

Heat demand constraint:

$$P_{T,out}^t \le H_r^t \tag{37}$$

Electric to heat constraint:

$$P_c^t = N_c(H_r^t - P_{T,out}^t) \tag{38}$$

TABLE 4 | The degree of membership and non-membership of conditions and decision attributes.

Data	Conditional attributes							
	P_1					P_2	P_3	P_4
	EL	L	M	H	EH	EN/ES/. . ./WN	F/G/P	H/L/M
1	(0.056,0.934)	(0.94,0.05)	(0,0)	(0,0)	(0,0)	NNE	F	H
2	(0,0)	(0.02,0.97)	(0.98,0.01)	(0,0)	(0,0)	ENE	F	H
3	(0,0)	(0,0)	(1,0)	(0,0)	(0,0)	E	F	H
4	(0,0)	(1,0)	(0,0)	(0,0)	(0,0)	WNW	F	L
5	(1,0)	(0,0)	(0,0)	(0,0)	(0,0)	W	F	H
6	(0,0)	(0.81,0.18)	(0.19,0.8)	(0,0)	(0,0)	NNW	G	L
7	(0,0)	(0,0)	(0.1,0.89)	(0.9,0.09)	(0,0)	WSW	G	L
8	(0,0)	(0,0)	(0.81,0.18)	(0.19,0.8)	(0,0)	E	P	M
. . .								
98	(0,0)	(0,0)	(0,0)	(0.9,0.09)	(0.1,0.89)	ESE	F	H
99	(0,0)	(0,0)	(0,0)	(0,0)	(1,0)	S	G	L
100	(0,0)	(0,0)	(0,0)	(0.12,0.87)	(0.88,0.11)	WSW	P	M

Data	Conditional attributes									
	P_5					P_6				P_7
	EH	H	M	L	EL	EL	L	M	H	EH	D/J/F
1	(0,0)	(0,0)	(0,0)	(0,0)	(1,0)	(0,0)	(1,0)	(0,0)	(0,0)	(0,0)	D
2	(0,0)	(0,0)	(0,0)	(0.7,0.29)	(0.3,0.69)	(0,0)	(0,0)	(1,0)	(0,0)	(0,0)	J
3	(0,0)	(0,0)	(0,0)	(1,0)	(0,0)	(0,0)	(0,0)	(1,0)	(0,0)	(0,0)	F
4	(0,0)	(0.11,0.88)	(0.89,0.1)	(0,0)	(0,0)	(0,0)	(0,0)	(0.26,0.73)	(0.74,0.25)	(0,0)	F
5	(0,0)	(0,0)	(1,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0.7,0.29)	(0.3,0.69)	(0,0)	J
6	(0,0)	(0,0)	(1,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0.92,0.07)	(0.08,0.91)	(0,0)	J
7	(0,0)	(0,0)	(0,0)	(0.2,0.79)	(0.8,0.19)	(0,0)	(0,0)	(1,0)	(0,0)	(0,0)	D
8	(0,0)	(0,0)	(0.34,0.65)	(0.66,0.33)	(0,0)	(0,0)	(0,0)	(1,0)	(0,0)	(0,0)	J
. . .											
98	(0,0)	(0,0)	(0,0)	(0,0)	(1,0)	(0,0)	(1,0)	(0,0)	(0,0)	(0,0)	J
99	(0,0)	(0,0)	(0,0)	(0,0)	(1,0)	(1,0)	(0,0)	(0,0)	(0,0)	(0,0)	D
100	(0,0)	(0,0)	(0,0)	(0,0)	(1,0)	(0.14,0.85)	(0.86,0.13)	(0,0)	(0,0)	(0,0)	D

Data	Decision attributes						
	Q_1				Q_2		
	VU	MU	GSU	GS+VS	GSU	MS	VS
1	(0,0)	(0,0)	(0,0)	(1,0)	(0,0)	(0.21,0.59)	(0.79,0.01)
2	(0,0)	(0,0)	(0,0)	(1,0)	(0,0)	(0.24,0.56)	(0.76,0.04)
3	(0,0)	(0,0)	(0,0)	(1,0)	(0,0)	(0.55,0.25)	(0.45,0.35)
4	(0,0)	(0,0)	(0,0)	(1,0)	(0,0)	(0.18,0.62)	(0.82,0)
5	(0,0)	(0,0)	(0.24,0.56)	(0.76,0.14)	(0.24,0.66)	(0.76,0.04)	(0,0)
6	(0,0)	(0,0)	(0,0)	(1,0)	(0,0)	(0.43,0.37)	(0.57,0.23)
7	(0,0)	(0.69,0.11)	(0.31,0.49)	(0,0)	(0.71,0.19)	(0,0)	(0,0)
8	(0,0)	(0,0)	(0,0)	(1,0)	(0,0)	(0.28,0.52)	(0.72,0.08)
. . .							
98	(0,0)	(0,0)	(0,0)	(1,0)	(0,0)	(0.27,0.53)	(0.73,0.07)
99	(0,0)	(1,0)	(0,0)	(0,0)	(0.79,0.11)	(0,0)	(0,0)
100	(0,0)	(0,0)	(0.16,0.64)	(0.84,0.06)	(0.18,0.72)	(0.84,0)	(0,0)

TABLE 5 | Decision rule attributes table.

Data	Conditional attributes							Decision attributes	
	P_1		P_2	P_3	P_5		P_7	Q	
1	EL (0.056,0.934)	L (0.94,0.05)	*	F	EL		*	MS (0.21,0.69)	VS (0.79,0.11)
2	L (0.02,0.97)	M (0.98,0.01)	*	F	*		J	MS (0.24,0.66)	VS (0.76,0.14)
3	M		*	*	c		F	MS (0.55,0.35)	VS (0.45,0.45)
5	EL		W	*	M		*	GSU (0.24,0.66)	MS (0.76,0.14)
7	M (0.1,0.89)	H (0.9,0.09)	*	G	L (0.2,0.79)	EL (0.8,0.19)	*	MU (0.69,0.21)	GSU (0.31,0.59)
8	M (0.81,0.18)	H (0.19,0.8)	E	P	*		J	MS (0.28,0.62)	VS (0.72,0.18)
					...				
99	EH		S	G	EL		*	MU	
100	H (0.12,0.87)	EH (0.88,0.11)	*	P	EL		*	GSU (0.18,0.72)	MS (0.82,0.08)

"" can be any value.*

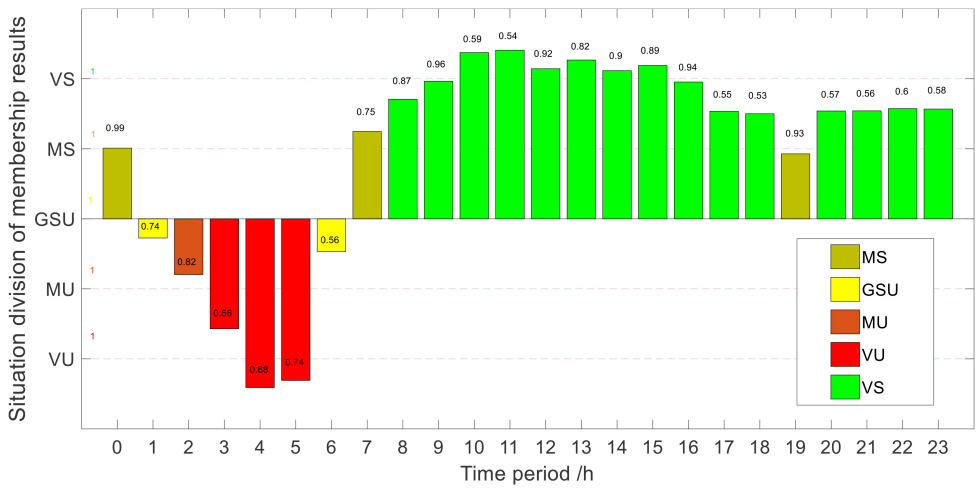

FIGURE 11 | The division result of consumption states of the electricity heat hydrogen system (EHHS) in mode 1.

Capacity constraint of TSU:

$$0 \le R_{cr}^{t} \le R_{cr,\max} \tag{39}$$

Where, R_{cr}^{t} is the heat storage capacity of the TSU in period t. $R_{cr,\max}$ is the maximum heat storage capacity of the TSU.

Power constraints of TSU:

$$P_{T,dg}^{t} \cdot \eta_{dg} \cdot \eta_{sr,T} \cdot \Delta t - P_{sc,T}^{t} / \eta_{sc,T} \cdot \Delta t \le R_{cr,\max} - R_{cr}^{t-1} \tag{40}$$

$$0 \le P_{T,out}^{t} \cdot \Delta t \le R_{cr}^{t} \tag{41}$$

Where, $P_{T,dg}^{t}$ is the wind power consumed by the EBU.

Constraints of thermal power units and cogeneration units are same as constraints in VS.

Optimization Model of MU System

Optimization objective function

In this state, the wind power needs to be consumed by the EBU and the EHU. The optimizing target is maximizing the output of the EBU.

$$\max E_{dg} = \sum_{t \in T_a} \Delta t P_{dg}^{t} \tag{42}$$

Where, T_a represents the number of scheduling periods in MU. E_{dg} is the wind power in MU. P_{dg}^{t} is the wind power consumed by the EBU.

Constraints

It is same as the power balance constraints in GSU.

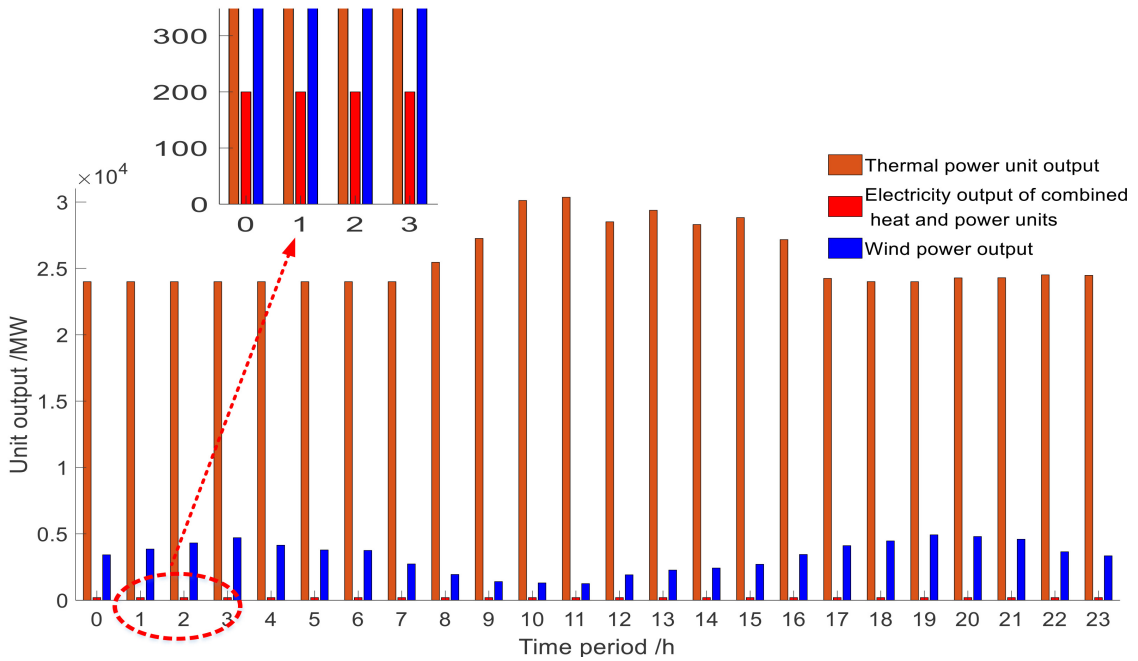

FIGURE 12 | Units output in mode 1.

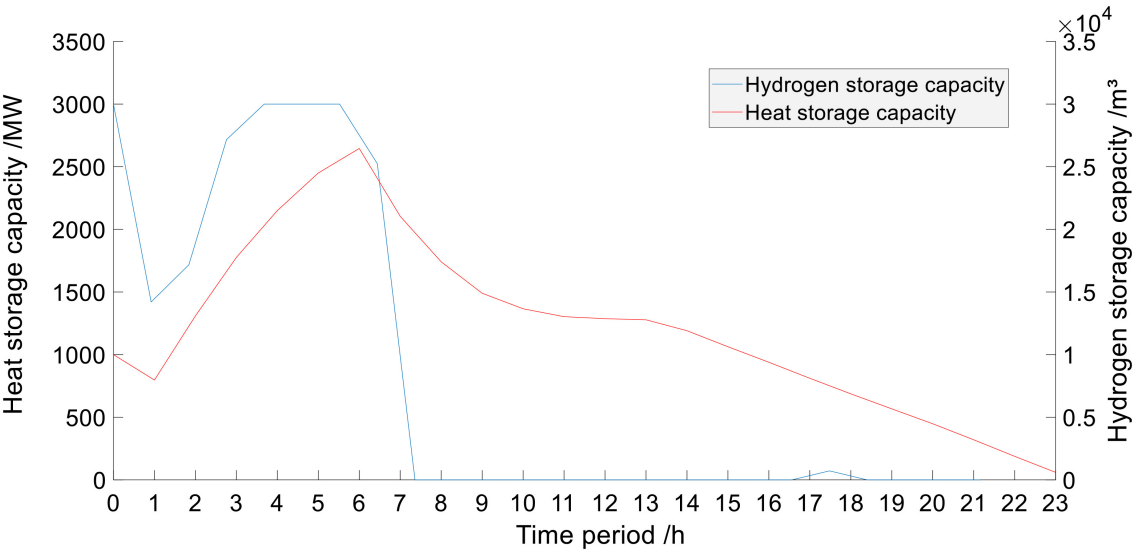

FIGURE 13 | Changes in capacity of the hydrogen storage unit (HSU) and thermal storage unit (TSU).

In addition to considering the equipment constraints in the previous consumption state, it is also necessary to restrict the capacity of the HSU.

$$0 \le R_{cq}^{t} \le R_{cq,\max} \tag{43}$$

$$R_{cq}^{t} = R_{cq}^{t-1} + P_{T,dj}^{t} \cdot \eta_{dj} \cdot \Delta t \cdot \eta_{sr,H} \tag{44}$$

Where, $P_{T,dj}^{t}$ is the wind power consumed by the EHU.

Optimization Model of VU System

Optimization objective function

In this state, the wind power cannot be completely consumed by the EBU and the EHU. The optimization goal is maximizing the wind power consumption.

$$\max E_f = \sum_{t \in T_b} \Delta t P_{fs}^{t} \tag{45}$$

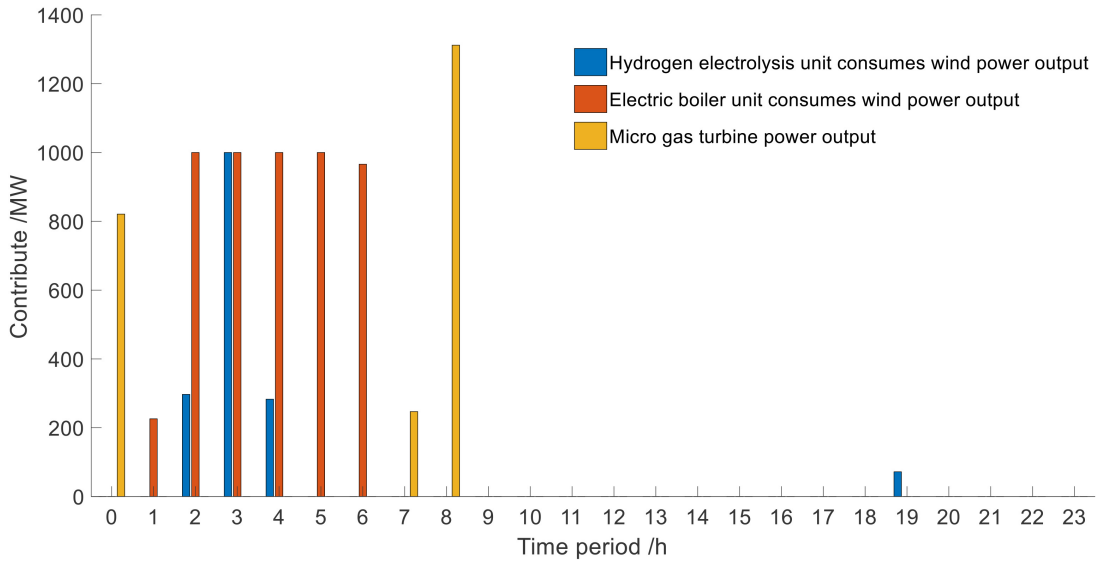

FIGURE 14 | Consumption and output of the electrolytic hydrogen unit (EHU), microturbine cogeneration unit (MCU), and electric boiler unit (EBU).

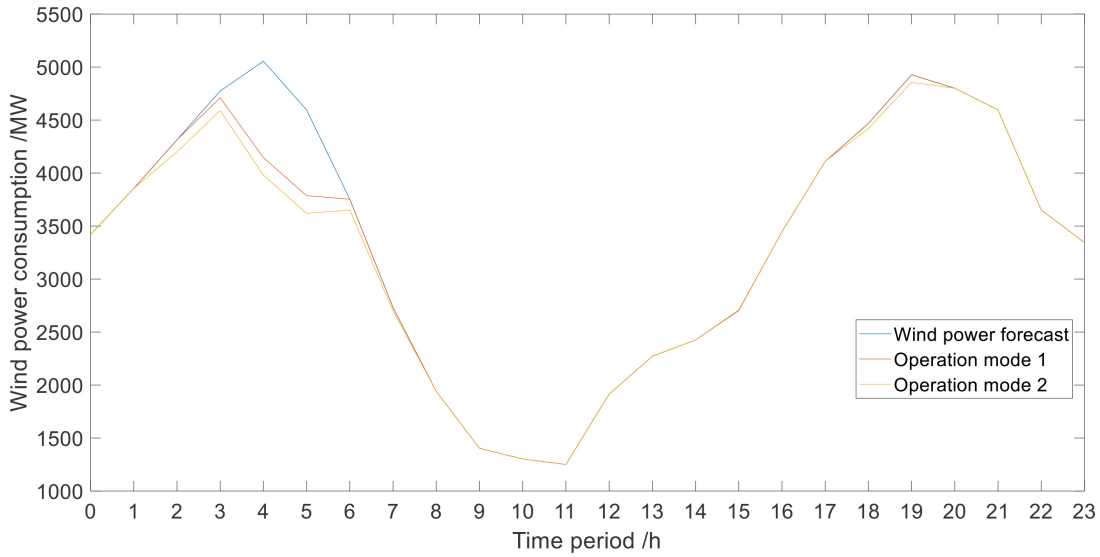

FIGURE 15 | Wind power consumption results under different modes.

Where, T_b represents the number of scheduling periods in VU. P_{fs} is the wind power actually consumed in VU.

Constraints

Power balance constraints:

$$\sum_{i}^{I_h} P_{h,i}^t + P_{f,s}^t + \sum_{i}^{I_c} P_{c,i}^t = P_d^t + P_{E,in}^t \tag{46}$$

$$H_c^t + P_{T,out}^t = H_r^t \tag{47}$$

The equipment constraints and the HHES constraints are same as constraints in MU.

Power System Optimal Dispatching Process Based on State Division Method

The optimal scheduling process of new energy consumption based on intuitionistic fuzzy rough set theory is shown in **Figure 7**.

CASE ANALYSIS

Instance Description

To highlight the advantages in solving the problem of wind power consumption, the measured condition attributes data of a

provincial power grid are taken as an example. The IEEE-30 bus system is used as the network structure of the EHHS. The system structure diagram is shown in **Figure 8**.

There are four thermal power units, two cogeneration units and two wind turbines being used for simulation. The total installed capacity of the wind farm is 5500 MW, and the installed capacity of each unit is shown in **Table 2**.

The heat-to-electricity ratio of the cogeneration unit is 1. The electro-hydrogen conversion coefficient of the EHU is 0.92, and the maximum installed power of the EHU is 1000 MW. The hydrogen storage and release efficiency of the HSU is 0.97, and the maximum installed capacity of the HSU is 30,000 m^3. The hydrogen-to-electric conversion efficiency of the MCU is 60%. The electric-to-heat conversion efficiency of the EBU is 0.96, and the maximum installed power of the EBU is 1,000 MW. The heat storage and release efficiency of the TSU is 0.9, and the maximum installed capacity of the TSU is 3,000 MWh. The selected scheduling period is 1 h, and one day's data are used for analysis. Assume that a typical day has stored heat of 1,000 MWh and hydrogen storage of 30,000 m^3.

A typical daily load curve and wind power output curve are shown in **Figure 9**. The heat load curve is shown in **Figure 10**.

Ignoring the line losses, equipment losses, and the effect of thermal network constraints, the system is simulated and analyzed in the following two operating modes:

Mode 1: Based on the state division method studied in this article, the HHES is introduced to participate in the scheduling, and the hybrid particle swarm optimization algorithm (Jiang and Ma, 2020; Liu et al., 2020; Lu and Song, 2020) is used for optimization.

Mode 2: Based on the three-state division method of the power system, electric energy storages and electric boilers were introduced to participate in dispatching and optimizing the system.

Example Demonstration

Consumption states of mode 1 are divided as follows. The consumption states division method in mode 2 is based on (Ge et al., 2019).

Related Consumption Data

The consumption data in a certain period of winter is selected to construct the knowledge base. **Table 3** is the part of the sample set, where P is the condition attribute, Q is the decision attribute. The value of the decision attributes in **Table 3** is the data of unoptimized scenario.

Intuitionistic Fuzzification of Sample Data

In the process of intuitionistic fuzzification of conditional attributes, five linguistic variables EL, L, M, H, and EH are set for the continuous attributes such as wind speed, outdoor temperature and room temperature. For wind direction attributes, 16 linguistic variables are used, such as North (N), North Northeast (NNE), Northeast (NE), East Northeast (ENE), East (E), East Southeast (ESE), Southeast (SE), South Southeast (SSE), South (S), South Southwest (SSW), Southwest (SW), West Southwest (WSW), West (W), West Northwest (WNW), Northwest (WN), and Northwest (NNW). For the electricity consumption period, they are divided into three periods, such as peak period (F), valley period (G), and flat period (P). For electricity price attributes, they are divided into three levels of high (H), medium (M), and low (L). For the monthly attribute, because this article only considers the consumption in the winter period, the data are divided into three levels: December (D), January (J), and February (F). Since the historical data of the above conditional attributes is sufficient, the intuition index of condition attributes and decision attributes can be 0.01 and 0.1.

If the result can be directly determined by η, then δ is not considered. Otherwise, the above two decision attributes will be considered to obtain the result. The results of intuitionistic fuzzy division are shown in **Table 4**.

Construction and Law of Decision Rule Set

For the membership results of decision attributes in the decision rule, it is first determined whether Q1 can get the result, and a conclusion will be drawn. Otherwise, the state membership result will be determined through Q2. The decision table obtained after attributes reduction and values reduction is shown in **Table 5**.

After obtaining the reduced decision rule table, the first and second rules are taken as an example: the first rule shows that when the system is in the peak period of electricity consumption with extremely low temperature and low wind speed, the system consumption state is VS. The second rule shows that during the peak electricity consumption period in January when the wind speed is moderate, the system consumption state is VS.

Matching the intuitionally fuzzy forecast data with the decision rule set, and considering the capacity of energy storage units to further determined the result, the state of the EHHS can be divided.

Result Analysis

In mode 1, the result of the system consumption state is shown in **Figure 11**.

In mode 1, when the state is MS, compared with the previous division method, the division method can more accurately determine the system optimization method. Take 7 and 19 as an example, the membership degrees of these two periods belong to MS. When the scheduling period is 7, the membership degree of MS is 0.75 and is closer to VS., so the MCU is used to provide power. At 19 o'clock, the optimized method is using EHU to optimize the system.

In addition to the advantage of the above division, at 2 o'clock, mode 2 determines that the system is in an emergency state and the electric boiler is used to help consume wind power. While in mode 1, combined with intuitive judgment, the system consumption state is MU. From the perspective of the optimized results after division, the optimization of the system in mode 1 can meet the demand for complete consumption, while there is a certain amount of abandonment in mode 2. It can be seen that under the new division method, not only the consumption state of the system can be divided more comprehensively, but also it can be judged more accurately.

Dividing the system consumption state based on the intuitionistic fuzzy rough set theory can comprehensively divide

the system consumption state, and the optimization method of the current system can be accurately selected.

In mode 1, the output of each unit in the system is shown in **Figure 12**.

In **Figure 12**, when the system is in a state that is less suitable for consumption, in order to increase the wind power consumption, the output of the conventional units is usually reduced. When in a state that is more suitable for consumption, the system can consume all the wind power. Because the output of the thermal power units is not reduced to the minimum output, the system still has a certain consumption potential. In addition, due to the participation of the HHES, the TSU can supply the heat load, and the cogeneration unit output can be reduced, and the overall economy is improved.

In mode 1, the changes in capacity of the HSU and the TSU are shown in **Figure 13**. The consumption and output of the EHU, MCU, and the EBU are shown in **Figures 14**.

It can be seen from **Figures 13**, **14** that when the state is VS, the MCU is working, but its operation is affected by the capacity of the HSU. The heat load of the system can also be supplied by the heat in the TSU. When the capacity of the HSU and the TSU do not meet the demand for consumption, the wind power will be abandoned.

Under different modes, the wind power consumption curve is shown in **Figure 15**.

From the optimization results in the **Figure 15**, on the basis of the new energy consumption state division, the wind power consumption can be significantly increased. Because the consumption state is divided before the optimization, the system is more suitable for consumption. The state in VS, MS, and GSU can meet the demand for complete consumption. When in MU and VU, the consumption demand during these periods can also be improved. Under the new division method, the division is more accurate and the optimization goals in each state can be more accurately achieved.

CONCLUSION

Aiming at the problem of blunt and inaccurate state division in the past, a dynamic state division method based on intuitionistic fuzzy rough set theory is established. The HHES is introduced to coordinate and cooperate with each unit, and a staged optimization model based on the division of the consumption state is established. Through analysis and comparison, a series of advantages of this model are obtained.

(1) Because the application of the EHU, HSU, and TSU in the EHHS has good economic and environmental protection, it has broad prospects in the application and development in the future energy market.
(2) Using intuitionistic fuzzy rough set theory to determine the system state, which can determine the consumption state more flexibly and accurately, it has obvious advantages in dealing with uncertain boundaries between various states.
(3) On the basis of dynamic division, the optimization effect is more obvious, and it has advantages of improving system operation economy and increasing wind power consumption.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author/s.

AUTHOR CONTRIBUTIONS

XL conceived the idea for the manuscript and wrote the manuscript with input from XZ and WZ. All authors have read and agreed to the published version of the manuscript.

FUNDING

This work was supported by the National Key R&D Program of China (2018YFA0702200) and Fundamental Research Funds for the Central Universities (N2004013).

REFERENCES

Chen, X., Rong, S., Guan, W., Hao, W., and Xu, M. (2020). Multi-source coordinated dispatching strategy for promoting wind power consumption. *Heilongjiang Electr. Power* 42, 480–486. doi: 10.13625/j.cnki.hljep.2020.06.003

Chen, Y., Xu, J., Shi, J., and Liu, Y. (2020). Three-branch decision model based on intuitionistic hesitation fuzzy set and its application. *Comput. Sci.* 47, 144–150.

Diao, H., Li, P., Wang, J., Wang, D., and Li, Z. (2020). Optimal scheduling of integrated energy system considering complementary coordination of electric and thermal energy storage. *Trans. China Electrotechnical Soc.* 35, 4532–4543. doi: 10.19595/j.cnki.1000-6753.tces.191340

Ge, W., Zhang, Y., Gao, C., Gao, K., Liu, X., and Ma, G. (2019). The phased optimization strategy of source-load storage system based on situation division of wind power consumption capacity. *Autom. Electr. Power Syst.* 43, 26–33.

Jiang, Y., and Ma, Y. (2020). Application of hybrid particle swarm and ant colony optimization algorithms to obtain the optimum homomorphic wavelet image fusion: introduction. *Ann. Transl. Med.* 8:1482. doi: 10.21037/ATM-20-5997

Kumar, A., and Prasad, P. S. V. S. S. (2020). Scalable fuzzy rough set reduct computation using fuzzy min-max neural network preprocessing. *IEEE Trans. Fuzzy Syst.* 28, 953–964. doi: 10.1109/TFUZZ.2020.2965899

Li, H., Yu, G., Tian, C., and Shang, G. (2020). An emergency decision making method based on fuzzy language TOPSIS. *J. Univ. Chin. Acad. Sci.* 37, 814–819.

Liu, W., Wen, J., Xie, C., Wang, W., and Liang, C. (2015). Power system source-load coordination multi-objective optimization method considering wind power consumption. *Proc. Chin. Soc. Electri. Eng.* 35, 1079–1088. doi: 10.13334/j.0258-8013.pcsee.2015.05.008

Liu, Z., Qin, Z., Zhu, P., and Li, H. (2020). An adaptive switchover hybrid particle swarm optimization algorithm with local search strategy for constrained optimization problems. *Eng. Appl. Artif. Intell.* 95:103771. doi: 10.1016/j.engappai.2020.103771

Lu, X., and Song, W. (2020). Research on application of hybrid particle swarm algorithm in distribution network optimization. *Small Microcomput. Syst.* 41, 1788–1792.

Pawlak, Z. (1982). Rough sets. *Int. J. Parallel Program.* 11, 341–356.

Ren, X., and Xue, F. (2018). A heuristic attribute reduction algorithm based on fuzzy neighborhood rough set. *Comput. Eng. Appl.* 54, 47–53.

Shu, Y., Zhigang, Z., Guo, J., and Zhang, Z. (2017). Analysis of key factors and solutions of new energy consumption. *Proc. Chin. Soc. Electri. Eng.* 37, 1–8. doi: 10.13334/j.0258-8013.pcsee.162555

Sun, H. (2021). Hydrogen energy is arousing great attention all over the world. *Int. J. Hydrog. Energy* 46, 2845–2846. doi: 10.1016/J.IJHYDENE.2020.08.212

Sun, H., Zhang, W., and Lei, Z. (2019). "Research on wind power consumption dispatching based on improved whale optimization algorithm," in *Proceedings of the International Conference on Intelligent Automation*, (Berlin: Springer), 506–514.

Teng, Y., Sun, P., Luo, H., and Chen, Z. (2019a). Autonomous optimization operation model of multi-source microgrid considering electric and thermal hybrid energy storage. *Proc. Chin. Soc. Electri. Eng.* 39, 5316–5324. doi: 10.13334/j.0258-8013.pcsee.181922

Teng, Y., Wang, Z., Li, Y., Ma, Q., Hui, Q., and Li, S. (2019b). Multi-energy storage system model based on electricity heat and hydrogen coordinated optimization for power grid flexibility. *CSEE J. Power Energy Syst.* 5, 266–274. doi: 10.17775/CSEEJPES.2019.00190

Wei, D., Xu, D., and Zhang, Y. (2020). A fuzzy evidential reasoning-based approach for risk assessment of deep foundation pit. *Tunn. Undergr. Space Technol.* 97:103232. doi: 10.1016/j.tust.2019.103232

Wu, D., Cheng, H., Zhao, J., Geng, J., and Hou, W. (2019). Wind power segmentation and thermal power peaking compensation method based on balancing cost. *Autom. Electr. Power Syst.* 43, 116–123.

Wu, Y., Guo, W., Tang, J., and Ren, Y. (2019). Summary of research on attribute reduction of several types of extended rough set models. *J. Yibin Univ.* 19, 29–38. doi: 10.19504/j.cnki.issn1671-5365.20190531.001

Xi, L., Wu, J., Xu, Y., and Sun, H. (2020). Automatic generation control based on multiple neural networks with actor-critic strategy. *IEEE Trans. Neural Netw. Learn. syst.* 99, 1–11. doi: 10.1109/TNNLS.2020.3006080

Xi, L., Yu, L., Xu, Y., Wang, S., and Chen, X. (2019). A novel multi-agent ddqn-ad method-based distributed strategy for automatic generation control of integrated energy systems. *IEEE Trans. Sustain. Energy* 11, 2417–2426. doi: 10.1109/TSTE.2019.2958361

Yang, J., Zhang, N., Wang, Y., and Kang, C. (2018). Multi-energy systems for renewable energy consumption: review and prospects. *Autom. Electr. Power Syst.* 42, 11–24.

Yang, Q., Zhi, C., Yuan, D., and Yang, H. (2020). Low carbon economic dispatching of wind power consumption based on coordinated heating of power off boiler and heat storage device. *Acta Energiae Solaris Sinica* 41, 21–28.

Zhan, J., and Sun, B. (2020). Covering-based intuitionistic fuzzy rough sets and applications in multi-attribute decision-making. *Artif. Intell. Rev.* 53, 671–701. doi: 10.1007/s10462-018-9674-7

Zhang, L., Zhan, J., Xu, Z., and Alcantud, J. C. R. (2019). Covering-based general multigranulation intuitionistic fuzzy rough sets and corresponding applications to multi-attribute group decision-making. *Inf. Sci.* 494, 114–140. doi: 10.1016/j.ins.2019.04.054

Zhang, P., Chen, X., and Fan, C. (2020). Research on a safety assessment method for leakage in a heavy oil gathering pipeline. *Energies* 13:1340. doi: 10.3390/en13061340

Zhang, R., Jiang, T., Li, G., Chen, H., Li, X., and Ning, R. (2018). Double-layer optimal dispatching of electricity-gas integrated energy system considering wind power conversion from electricity to gas. *Proc. Chin. Soc. Electri. Eng.* 38, 5668–5678. doi: 10.13334/j.0258-8013.pcsee.172310

Zhang, Y., Lu, Q., Zhang, N., Wang, H., Liu, L., and Sun, H. (2020). Research on coordinated operation of electric-heat peak shaving resources for wind power consumption. *Power Syst. Technol.* 44, 1350–1362. doi: 10.13335/j.1000-3673.pst.2019.1796

Zhou, S., Dai, S., Xu, D., Lang, Y., and Hu, L. (2020). Combined electric and heating dispatch model considering source-side flexible transformation and adjustable electric and heating load. *Power Syst. Technol.* 44, 2254–2262. doi: 10.13335/j.1000-3673.pst.2019.1139

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Energy Research

ORIGINAL RESEARCH
published: 22 June 2021
doi: 10.3389/fenrg.2021.693878

Study on a Frequency Fluctuation Attenuation Method for the Parallel Multi-VSG System

Zhenao Sun, Fanglin Zhu and Xingchen Cao*

College of Information Science and Engineering, Northeastern University, Shenyang, China

Virtual synchronous generator (VSG) is one of the inverter control methods which can provide extra virtual moment of inertia and achieve frequency support by mimicking the output characteristics of a rotating synchronous generator (RSG), which makes VSG particularly suitable for multi-access point applications, or called the multi-VSG system. However, frequency fluctuations may often occur in the case of power variation. In terms of this issue, this study presents the small-signal state-space model of VSG and analyzes the cause of frequency fluctuation, first. And then, a novel VSG control method is proposed for frequency fluctuation attenuation. The proposed method is especially fit for the parallel multi-VSG system because it does not take grid angular frequency into computation so that it can get rid of the ill effect introduced through PLL. The damping power item is reconstrued with a new means to judge whether the system is in a steady state or not. At the same time, the parametric design method of the proposed method is also figured out. Finally, the simulation experiments are performed, and the results verify that the proposed method performs better than the conventional one in terms of dynamic response and power-sharing among the multi-VSG system.

Keywords: parallel multi-VSG system, frequency fluctuation, grid angular frequency, damping power item, dynamic performance

OPEN ACCESS

Edited by:
Yonghao Gui,
Aalborg University, Denmark

Reviewed by:
Kenneth E. Okedu,
National University of Science and Technology (Muscat), Oman
Krishnakumar R. Vasudevan,
Universiti Tenaga Nasional, Malaysia

****Correspondence:***
Zhenao Sun
sunzhenao@mail.neu.edu.cn

Specialty section:
This article was submitted to Smart Grids, a section of the journal Frontiers in Energy Research

Received: *12 April 2021*
Accepted: *01 June 2021*
Published: *22 June 2021*

Citation:
Sun Z, Zhu F and Cao X (2021) Study on a Frequency Fluctuation Attenuation Method for the Parallel Multi-VSG System. Front. Energy Res. 9:693878. doi: 10.3389/fenrg.2021.693878

INTRODUCTION

In recent years, with the development of distributed energy resources (DERs) and microgrids (DG), more and more power electronic devices have been added to the grid. However, huge challenges, such as the problem of lack of inertia, damping, and rotational reserve capacity, have emerged with a large number of power electronic equipment. In order to solve these issues, virtual synchronous generator (VSG) has been employed to the control system of distributed energy power generation equipment (Hafner et al., 2011; Mo et al., 2017; Wang et al., 2020), which can make the equipment have frequency and voltage regulation characteristics similar to conventional synchronous generators.

The basic theory of VSG has been explained in the studies by D'Arco and Suul (2014)and Driesen and Visscher (2008), and the mathematical model and implementation are investigated under the title of synchronverter (Zhong and Weiss, 2011; Zhong et al., 2014). VSG strategy has been introduced in frequency, voltage, and active and reactive power flow control (Wang et al., 2020). Consequently, VSG control strategy has been applied to different devices and occasions, such as energy storage (Ma et al., 2017), doubly fed induction generators (Hwang et al., 2017), high-voltage

direct current transmission (Aouini et al., 2016), and direct current (DC) generators (Wu et al., 2017).

Unfortunately, in the VSG control system, when the distributed power supply fluctuated drastically, the power and frequency outputs of the system are prone to oscillations. In order to effectively eliminate the effects of fluctuations on the system, various control methods have been proposed (Wu et al., 2016; Li et al., 2017a; Alsiraji and El-Shatshat, 2017; Shi et al., 2018; Wang et al., 2018a). In the study by Wu et al. (2016), by adjusting the damping coefficient, moment of inertia, and other related parameters of the VSG, the output characteristics can be directly changed to achieve suppression of power and frequency fluctuations. A bang–bang control method with adaptive moment of inertia has been proposed (Shi et al., 2018) by selecting various rotational inertia under different operating conditions. In order to design the parameters more conveniently, a small-signal model has been introduced in the study byWang et al. (2018a) to determine the best damping coefficient and moment of inertia in the system. Moreover, a model that aims at the minimum transient response time has been established, and the influence of the frequency amplitude and rate of change has been considered (Li et al., 2017a). However, the damping coefficient of the above control method was set to zero under specific operating conditions, which led to a limited range of applications. Therefore, an optimal damping ratio control strategy has been proposed in the study by Alsiraji and El-Shatshat (2017) to improve the transient frequency, power response, and stability of the VSG system. Nevertheless, this method cannot maintain the best damping ratio during the entire operation of the system.

However, the transient condition tolerance of VSG units is much less than that of a real synchronous generator (Alipoor et al., 2015). To solve the issue of frequency fluctuation, at present, more and more scholars are devoted to the research of parameter adaptation. Damping and moment of inertia designed in the study by Li et al. (2017b) follow the law that damping is reduced in the interval of $\Delta\omega$ ($d\omega/dt$) > 0 and remains at D_{0in} in other times, while the moment of inertia setting rule is opposite. However, the optimal relationship and extreme values between virtual inertia and damping coefficient have not been studied in depth. Besides, in the initial state after the disturbance, virtual inertia and the damping coefficient are small, which greatly limit the transient stability. The study by Alipoor et al. (2018) does not include a theoretical analysis of the influence of the damping coefficient. In some specific cases, this control strategy is difficult to implement. Although the study by Wang et al. (2018b) contains a comprehensive theoretical analysis of damping and moment of inertia and optimal parameter selection, there are only two virtual inertia values in this control strategy, which reduces the operating performance of the controller during critical disturbances.

The methods mentioned above, which improve the oscillation of the VSG system frequency and active power, were mainly realized by changing the system parameters.

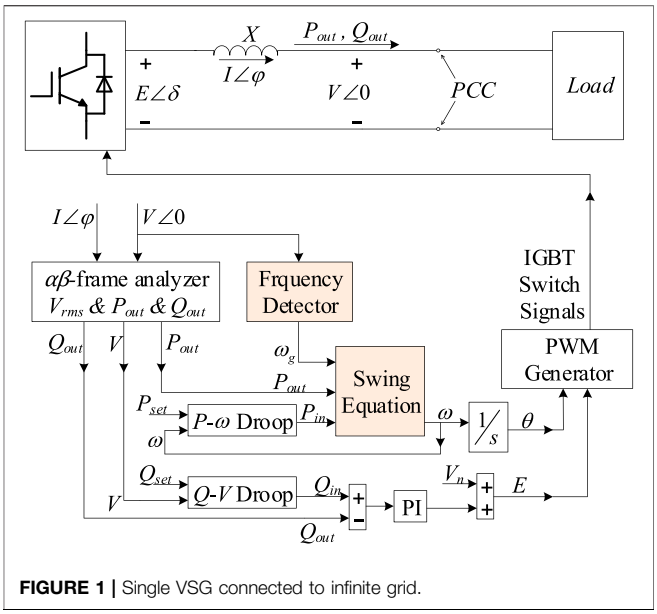

FIGURE 1 | Single VSG connected to infinite grid.

Traditional methods have limited the application of control strategies that change parameters. At the same time, these methods all aimed at the operation of a single VSG, failing to reflect the power and frequency changes of multiple VSGs operating in parallel. According to the analysis above, this study proposes a new control strategy to suppress the attenuation of the frequency and active power of multiple VSGs.

In summary, the control method of the parallel multi-VSG system should be different from that of the single VSG system. However, few of the existing methods can improve the transient stability according to the characteristics of the multi-VSG system. Some certain special means should be taken to enhance the transient performance of the former system. A frequency fluctuation attenuation method is needed for the parallel multi-VSG system. Thus, this study aimed to investigate the parallel operation of VSGs with a damping correction term added to the swing equation. The addition of the extra term intends to achieve better oscillation damping in a multi-VSG environment. The load changing transients and eigenvalue analysis were employed to validate the transient and small-signal stability of the system with the proposed strategy.

The article is arranged as follows. Section *Single VSG Connected to the Infinite Grid* presents the model of a single VSG connected to the infinite grid. Section *The Course of Frequency Fluctuation* analyzes the course of frequency fluctuation in the parallel multi-VSG system. Section *The Proposed Control Method for the Parallel Multi-VSG System* proposes a novel control method for frequency fluctuation attenuation in the parallel multi-VSG system. Section *Simulation Results* exhibits and analyzes the simulation results to verify that the proposed method performs better than the conventional one. At last, Section *Conclusion* summarizes the article conclusions.

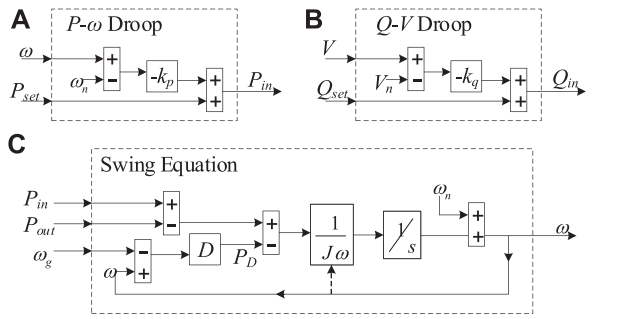

FIGURE 2 | Block diagram of **(A)** *P-ω* Droop, **(B)** *Q-V* Droop, and **(C)** swing equation.

SINGLE VIRTUAL SYNCHRONOUS GENERATOR CONNECTED TO THE INFINITE GRID

Two Control Loops of a Virtual Synchronous Generator

Figure 1. illustrates the model of a single VSG connected to the infinite grid, with the DC side of the VSG ignored. **Figure 1** shows two main control loops: active power control loop (APCL) and reactive power control loop (RPCL). The *P-ω* droop, *Q-V* droop, and swing equation are contented. Their detailed block diagrams are shown in **Figure 2**.

In **Figure 2**, k_p and k_q are the droop coefficients of each loop correspondingly, J is the virtual rotational inertia, D is the damping factor, ω_n is the nominal value of the grid angular frequency, and ω_0 is the grid angular frequency on the steady-state operating point.

Equations 1–4 describe the APCL:

$$P_{\text{in}} = P_{\text{set}} - k_p(\omega - \omega_n), \tag{1}$$

$$P_D = D \cdot (\omega - \omega_g), \tag{2}$$

$$P_{\text{in}} - P_{\text{out}} - P_D = J\omega\frac{\mathrm{d}\omega}{\mathrm{d}t} \approx J\omega_0\frac{\mathrm{d}\omega}{\mathrm{d}t} \Leftrightarrow \omega \approx \omega_n + \frac{1}{J\omega_n}\int (P_{\text{in}} - P_{\text{out}} - P_D)\mathrm{d}t, \tag{3}$$

$$\theta = \int \omega \mathrm{d}t \Leftrightarrow \frac{\mathrm{d}\theta}{\mathrm{d}t} = \omega, \tag{4}$$

and **Eqs 5, 6** describe the RPCL (Hafner et al., 2011; Mo et al., 2017):

$$Q_{\text{in}} = Q_{\text{set}} - k_q(V - V_n), \tag{5}$$

$$E = V_n + \frac{K_I}{s} \cdot (Q_{\text{in}} - Q_{\text{out}}). \tag{6}$$

Figure 3 is the vector diagram of **Figure 1** (ignore R). **Figure 3** reveals the role of δ, formulated as **Eqs 7–9**.

$$P_{\text{out}} = \frac{3EV\sin\delta}{X}, \tag{7}$$

$$Q_{\text{out}} = \frac{3(EV\cos\delta - V^2)}{X}, \tag{8}$$

$$\delta = \int (\omega - \omega_g)\mathrm{d}t \Leftrightarrow \frac{\mathrm{d}\delta}{\mathrm{d}t} = \omega - \omega_g.$$

It is noticed that APCL is key to VSG control because the main characteristic of the VSG is to introduce a virtual rotational inertia J in order to mimic the dynamic performance of an RSG, and APCL is more complex than RPCL.

Decoupling of the Two Control Loops

The small-signal model of **Eqs 7, 8** can be expressed in a matrix as **Eq. 9** (Zhong and Weiss, 2011):

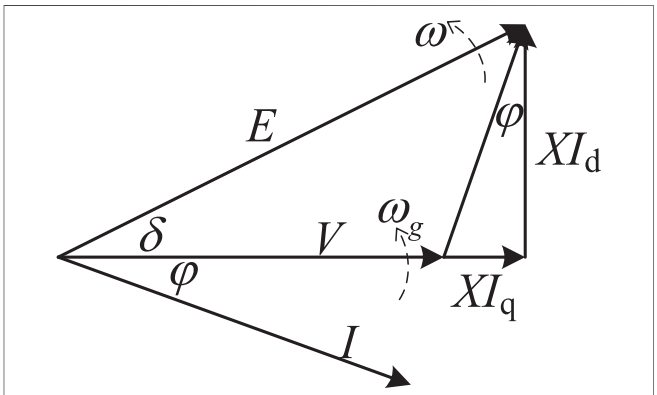

FIGURE 3 | Vector diagram of a VSG.

$$\begin{bmatrix}\Delta P_{\text{out}}\\ \Delta Q_{\text{out}}\end{bmatrix} = \frac{3}{X}\cdot M \cdot \begin{bmatrix}\Delta\delta\\ \Delta E\\ \Delta V\end{bmatrix}, \quad M = \begin{bmatrix} E_0V_0\cos\delta_0 & V_0\sin\delta_0 & E_0\sin\delta_0 \\ -E_0V_0\sin\delta_0 & V_0\cos\delta_0 & E_0\cos\delta_0 - 2V_0\end{bmatrix}. \tag{9}$$

It seems like the two control loops are coupling with each other, but we can do some approximations to decouple them and reveal the decoupling conditions.

In engineering practice, the short-circuit radio I_{SC}/I_n is usually designed more than 10, formulated as (Ma et al., 2017)

$$\frac{I_{SC}}{I_n} = \frac{E_n/X}{P_n/3V_n} = \frac{3E_nV_n}{P_nX} > 10. \tag{10}$$

Substituting **Eq. 6** into **Eq. 10**, we get **Eq. 11**:

$$\sin\delta_n = \frac{P_nX}{3E_nV_n} < 0.1. \tag{11}$$

In this case, δ is usually a tiny angle. For example, a 10-kW inverter connects to a 380/220 V three-phase grid with a 4-mH

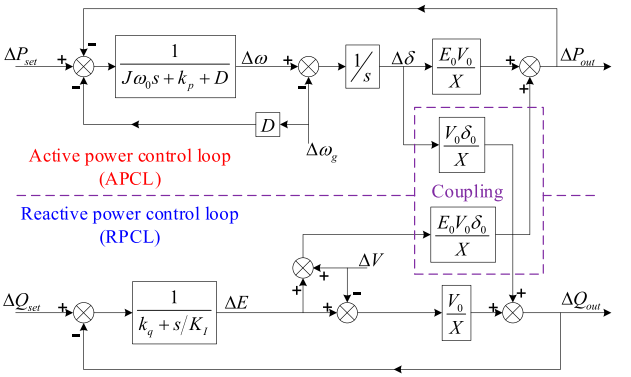

FIGURE 4 | Small-signal control model.

filter inductor: $\sin\delta = 0.0862 < 0.1$, then $\delta = 0.0863$ and $\tan\delta = 0.865$; they are approximately equal, and $\cos\delta = 0.9963 \approx 1$. So, in VSG control, we consider

$$\begin{cases} \sin\delta \approx \delta \approx \tan\delta \\ \cos\delta \approx 1 \end{cases}. \tag{12}$$

In another word,

$$E \approx V. \tag{13}$$

Substituting **Eqs 12**, **13** into **Eq. 10** and writing the small-signal model of **Eq. 9**, we get **Eq. 14**:

$$\begin{cases} \Delta P_{\text{out}} = \dfrac{3E_0V_0}{X}\cdot\Delta\delta + \dfrac{3V_0\delta_0}{X}\cdot(\Delta E + \Delta V) \\ \Delta Q_{\text{out}} = \dfrac{3E_0V_0\delta_0}{X}\cdot\Delta\delta + \dfrac{3V_0}{X}(\Delta E - \Delta V) \\ \Delta\delta = \dfrac{\Delta\omega - \Delta\omega_g}{s} \end{cases}. \tag{14}$$

Substituting **Eqs 1**, **2** into **Eq. 3** and deducing its small-signal model, we have **Eq. 17**:

$$\begin{aligned} J\omega_0\cdot s\Delta\omega &= \Delta P_{\text{set}} - k_p\Delta\omega - D\left(\Delta\omega - \Delta\omega_g\right) - \Delta P_{\text{out}} \\ \Rightarrow \Delta\omega &= \frac{\Delta P_{\text{set}} - \Delta P_{\text{out}} - D\cdot\Delta\omega_g}{J\omega_0 s + k_p + D}. \end{aligned} \tag{15}$$

Similarly, substituting **Eqs 5**, **13** into **Eq. 6** and deducing its small-signal model, we have **Eq. 16**:

$$\begin{aligned} \Delta E &= \text{PI}\cdot\left(\Delta Q_{\text{set}} - k_q\Delta E - \Delta Q_{\text{out}}\right) \\ \Rightarrow \Delta E &= \frac{\Delta Q_{\text{set}} - \Delta Q_{\text{out}}}{k_q + (s/K_I)}. \end{aligned} \tag{16}$$

Then, the small-signal model of **Figures 1**, **2** can be illustrated in **Figure 4**:

For $\Delta P_{\text{out}}(s)$, $\Delta\delta$ is the input signal and ΔE and ΔV are the disturbance input signals. If we can prove the loop gain from $\Delta\delta$ to $\Delta P_{\text{out}}(s)$ is much larger than that from ΔE+ΔV to $\Delta P_{out}(s)$, we can overlook the coupling item from RPCL to APCL, and thus consider that the APCL is decoupled from RPCL.

Overlooking the coupling items, define the forward path gain of APCL as $F_{fp}(s)$:

$$F_{fp}(s) = \frac{1}{J\omega_0 s + k_p + D}\cdot\frac{1}{s}\cdot\frac{3E_0V_0}{X}. \tag{17}$$

Then the transfer functions between $\Delta P_{\text{out}}(s)$ and $\Delta\delta(s)$ or $E(s)$+$V(s)$ can be formulated as (Alsiraji and El-Shatshat, 2017; Alipoor et al., 2015), separately.

$$F_1(s) = \frac{\Delta P_{\text{out}}(s)}{\Delta\delta(s)} = \frac{3E_0V_0/X}{1 + F_{pf}(s)}, \tag{18}$$

$$F_2(s) = \frac{\Delta P_{\text{out}}(s)}{\Delta E(s) + \Delta V(s)} = \frac{3V_0\delta_0/X}{1 + F_{pf}(s)}. \tag{19}$$

Since $|F_1(s)|/|F_2(s)| >> 1$, compared to $F_1(s)$, $F_2(s)$ can be neglected; in another word, APCL is decoupled from RPCL. In the similar way, RPCL can also be decoupled from APCL approximately.

So, we can analyze the two control loops separately, in the condition where short-circuit ratio is more than 10, which is always satisfied in the inverter design.

THE COURSE OF FREQUENCY FLUCTUATION

Intrinsic Oscillation Mode

The VSG has a better transient stability because it emulates the swing equation of an RSG to provide virtual rotational inertia and extra frequency support to the grid. However, the emulation of swing equation of a conventional synchronous generator (SG) also introduces an oscillatory mode, which makes VSG-controlled inverters subjected to intrinsic low-frequency oscillation.

Define K as the transient synchronizing power coefficient in **Eq. 22**:

$$K = \frac{\partial P_{\text{out}}}{\partial\delta} = \frac{3E_0V_0\cos\delta_0}{X} \approx \frac{3V_0^2}{X} \approx \frac{3V_n^2}{X}. \tag{20}$$

Arrange **Eqs 1–3**, **9**, **20** in the state-space form shown in **Eq. 21**:

$$\begin{cases} \dot{x} = Ax + Bu \\ y = x \end{cases}. \tag{21}$$

In **Eq. 23**, the output vector **y**, the state vector **x**, the input vector **u**, the state matrix **A**, and the input matrix **B** are as **Eq. 22**:

$$\begin{aligned} y = x &= \begin{bmatrix} \Delta\omega & \Delta P_{out} \end{bmatrix}^T \quad u = \begin{bmatrix} \Delta P_{set} & \Delta\omega_g \end{bmatrix}^T \\ A &= \begin{bmatrix} -\dfrac{k_p + D}{J\omega_0} & -\dfrac{1}{J\omega_0} \\ K & 0 \end{bmatrix} \quad B = \begin{bmatrix} \dfrac{1}{J\omega_0} & \dfrac{D}{J\omega_0} \\ 0 & -K \end{bmatrix}. \end{aligned} \tag{22}$$

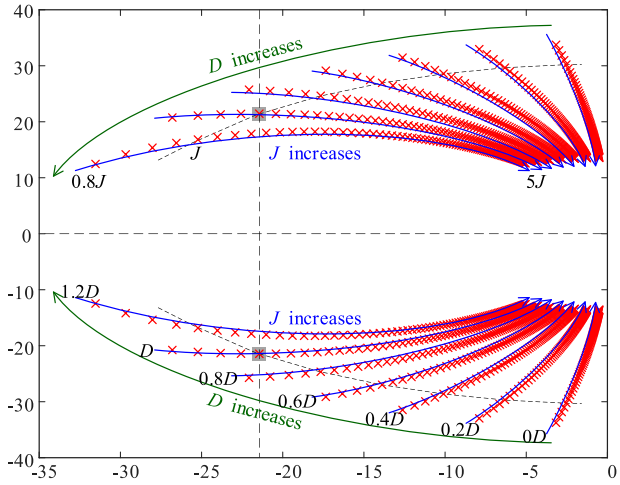

FIGURE 5 | Eigenvalue loci when J or D varies in the underdamping state.

Normally, the system operates in the underdamping state, and the eigenvalues of **A** can be deduced as **Eq. 23**:

$$\lambda_{1,2} = -\frac{k_p + D}{2J\omega_0} \pm j\frac{\sqrt{4J\omega_0 - (k_p + D)^2}}{2J\omega_0} \Leftrightarrow \lambda_{1,2} = \sqrt{\frac{K}{J\omega_0}} e^{j\left(\pi \pm \arccos\left((k_p+D)/(2\sqrt{KJ\omega_0})\right)\right)}. \tag{23}$$

Therefore, the undamped natural frequency ω_{un} and the damping ratio ζ of the intrinsic oscillation of the VSG can be represented as **Eq. 24**.

$$\omega_{\text{un}} = \sqrt{\frac{K}{J\omega_0}}, \quad \zeta = \frac{k_p + D}{2\sqrt{KJ\omega_0}}. \tag{24}$$

From **Eq. 24**, we can conclude that the undamped natural frequency ω_{un} will decrease as the virtual rotational inertia J increases, which is not influenced by the damping factor D. At the same time, larger J will result in smaller ζ, which may make the system more prone to oscillation. However, D can also influence ζ positively. Those relations can also be confirmed from the eigenvalue loci plot in **Figure 5**. The parameters involved are presented in **Table 1**. In **Figure 5**, J and D vary in proportion.

Frequency Fluctuation in the Parallel Multi-VSG System

The VSG mimics the output characteristics of the RSG, so the frequency fluctuation of the parallel multi-VSG system can also be compared to the state of the multi-RSG system. In the multi-RSG system, in order to keep in sync, the parallel RSG continuously adjusts the angular frequency of their respective rotors, which might cause a frequency fluctuation of the grid. In this situation, it is even more difficult to keep the RSGs in sync, resulting in continuous frequency fluctuation.

Similar situation can be seen in the parallel multi-VSG system. Because of the particularity of power electronic equipment, the controller of a VSG cannot obtain the grid angular frequency directly. A frequency detector, such as a phase-locked loop (PLL), is a must in the conventional VSG control method.

However, the response rate of digital PLL is much slower than that of mechanical PLL, which makes the frequency fluctuation in the parallel multi-VSG system even worse. Thus, a novel PLL-free VSG control method is needed for the frequency fluctuation attenuation in the parallel multi-VSG system.

THE PROPOSED CONTROL METHOD FOR THE PARALLEL MULTI-VSG SYSTEM

Novel Constitution of Damping Power Item P_D

The damping power item P_D is the only item that needs a frequency detector in the basic VSG control method. In order to get rid of the PLL, we must reconstruct the item P_D without using a frequency detector. So, we will first analyze the mechanism of original P_D in the basic VSG control, and then come up with a new way to obtain P_D.

From the study by Zhong et al. (2014), we can notice that P_D remains at 0 in a steady state and makes differences only in transient. The basic method takes the comparison of ω and ω_g as a way to assess whether the system is operating in a steady-state condition or not. If $\omega = \omega_g$, the system is operating in a steady-state condition and P_D should remain at 0. If $\omega \neq \omega_g$, the system is in the transient condition.

So that, if we want to obtain P_D in a novel way without using frequency detector, we must find a new way to assess the steady-state condition. In the system, P_{in} and P_{out} share a similar relationship with ω and ω_g. If $P_{\text{in}} = P_{\text{out}}$, the system is operating in the steady-state condition; if not, transient.

Thus, we can conclude that P_D should satisfy the following requirements:

(1) It should contain a factor, for example, $P_{\text{in}} - P_{\text{out}}$, to estimate whether the system is operating in the steady-state condition.
(2) It must turn to and remain at zero in the steady-state condition so that it will not affect the steady operating point of droop relation in the steady-state condition.

TABLE 1 | Same parameters involved.

Parameter	Value	Comments
S_n and S_{base}	10 kVA	
V_n and V_{base}	220 V	
$\omega_n \approx \omega_0$	100 π rad/s	f = 50 Hz
X	1.26 Ω	L = 4 mH, $K = 3V_n^2/X$ = 115,546
k_p	637	$k_p^* = k_p\omega_n/S_{\text{base}}$, k_p^* = 20 p.u.
J	0.4	$M^* = J\omega_0^2/S_{\text{base}}$, M^* = 4 s
D	4,752	ζ = 1/sqrt(2)

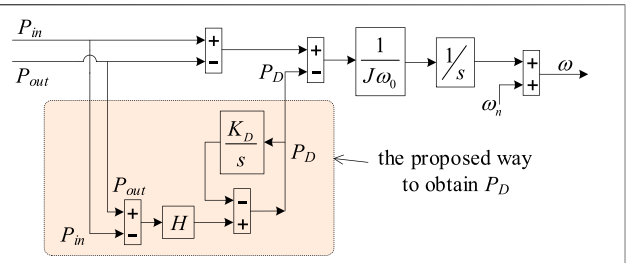

FIGURE 6 | Block diagram of the swing equation of the proposed method.

(3) Even if it introduces extra poles, the dominant poles of the new system should be closed to or coincident with the original two poles.
(4) It must contain a $D\Delta\omega$ item in its small-signal model so that the state matrix **A** and its eigenvalues will not change. Since P_{in} contains a $k_p{\cdot}\omega$ item and it will become a $k_p{\cdot}\Delta\omega$ item in its small-signal model, we do not need to constitute it dedicatedly and only need to adjust its coefficient.

To meet the four abovementioned requirements, we propose a novel way to obtain P_D, formulated as

$$P_D = H \cdot (P_{\text{out}} - P_{\text{in}}) - \frac{K_D}{s} \cdot P_D, \tag{25}$$

where H is the output power error amplification factor and K_D is the self-integral coefficient.

The first requirement is satisfied obviously. The second requirement is satisfied through the integral item, which will force P_D to turn to and remain at 0 in the condition of $K_D > 0$. The third and fourth requirements need proper parameter design, which will be detailed in the next subsection.

Figure 6 shows the block diagram of the swing equation of the proposed method using the novel way to obtain P_D. Comparing it with the original one shown in **Figure 2C**, it is noticed that the grid frequency does not participate in the computation; thus, it can avoid the ill effect of frequency detector.

Parameter Design of the Proposed Control Method

There are two parameters in **Eq. 25**, H and K_D, whose values need to be decided, and we can adjust these two parameters to meet the third and the fourth requirements.

To meet the third requirement, we should first deduce the eigenvalues of the proposed system. Arrange **Eqs 1**, **3**, **4**, **9**, **25** in the state-space form as shown in **Eq. 21**, and use a superscript to distinguish the corresponding vector or matrix in **Eq. 22**; then we can get **Eq. 26**.

In **Eq. 11**, most parameters in $\mathbf{A}'$ are deterministic, except for H and K_D, which correspond to ω_{un} and ζ. We need to find a combination of H and K_D, which will make the dominant poles of the new system closed to or to be coincident with the original ones, as shown in **Eq. 23** and **Figure 5**.

$$A' = \begin{bmatrix} -\left(k_p(1+H)\right)/J\omega_0 & (-(1+H)/J\omega_0) & K_D/J\omega_0 \\ K & 0 & 0 \\ Hk_p & H & -K_D \end{bmatrix}$$
$$B' = \begin{bmatrix} (1+H)/J\omega_0 & 0 \\ 0 & -K \\ -H & 0 \end{bmatrix} \quad y' = x' = \begin{bmatrix} \Delta\omega \\ \Delta P_{\text{out}} \\ P_D \end{bmatrix}^{\text{T}} = \begin{bmatrix} x \\ P_D \end{bmatrix}^{\text{T}}. \tag{26}$$

In order to get the eigenvalue expressions of $\mathbf{A}'$, we plot the eigenvalue loci with varying H and K_D to find a fitful combination, as shown in **Figure 7**. The eigenvalue loci are shown in **Figure 6**. With two parameters to be decided, this is something like "trial and error." But later in this section, we will prove that there is only one parameter that needs to be decided, and the root locus alike method can be used in this situation.

It is noticed that on the condition that $K_D = 180$ and $H = 7.4$, the dominant poles of the new system are almost coincident with the original poles, and the other pole is assigned much farther to the left than the dominant, which satisfies the third requirement.

What is more; substituting $H = 7.4$ into $\mathbf{A}'$, we notice that $D = H{\cdot}k_p$ and $\mathbf{A}'_{11} = \mathbf{A}_{11}$, which means the fourth requirement is also satisfied. Thus, we know that the conjugate eigenvalues are the dominant poles of $\mathbf{A}'$. If $\mathbf{A}'_{11} = \mathbf{A}_{11}$, then the dominant poles of $\mathbf{A}'$ will be the same as the original poles of **A**. So, we can use the equation $\mathbf{A}'_{11} = \mathbf{A}_{11}$ or $D = H{\cdot}k_p$ to obtain H directly, leaving K_D itself to be determined by the eigenvalue loci.

With only one single variable parameter to be determined, we can use a method similar to the root locus method to choose a suitable value on eigenvalue loci. The very single variable parameter of the mentioned method is K_D, not the open loop gain in the root locus method. And this method can be named as root locus alike method.

There are two parameters in the proposed method to be decided, H and K_D. H can be calculated directly, and K_D is picked up by the root locus alike method. No complicated formula derivation is involved in the parameter design approach, so this method can simplify the parameter design process.

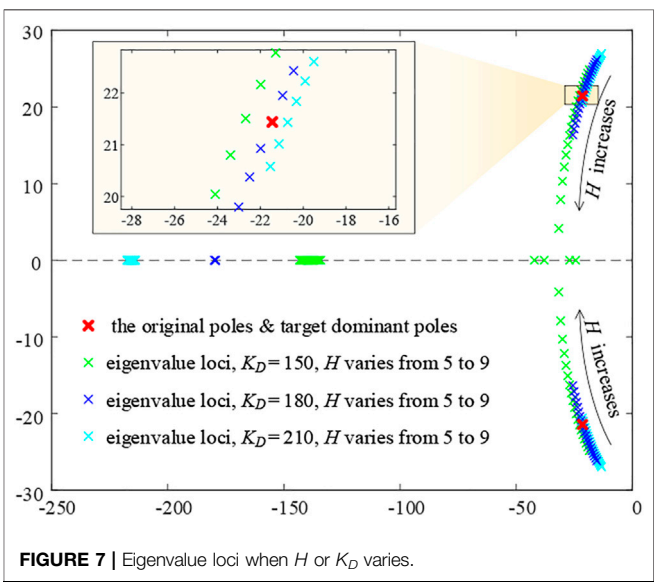

FIGURE 7 | Eigenvalue loci when H or K_D varies.

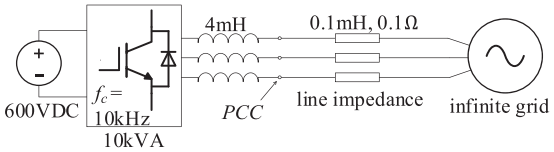

FIGURE 8 | Simulation topology of a VSG connected to the infinite grid.

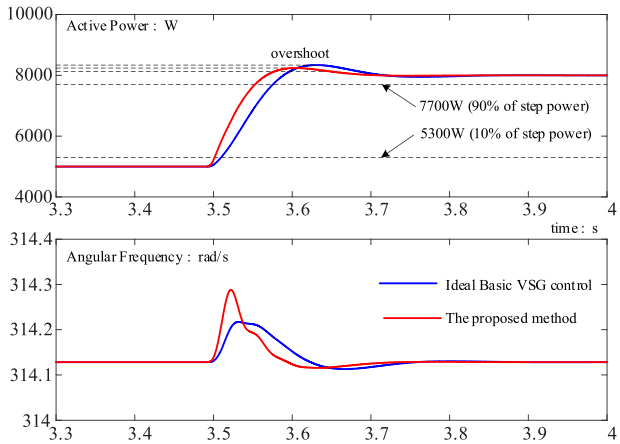

FIGURE 9 | Simulation results I: different VSG methods in the grid-connected mode.

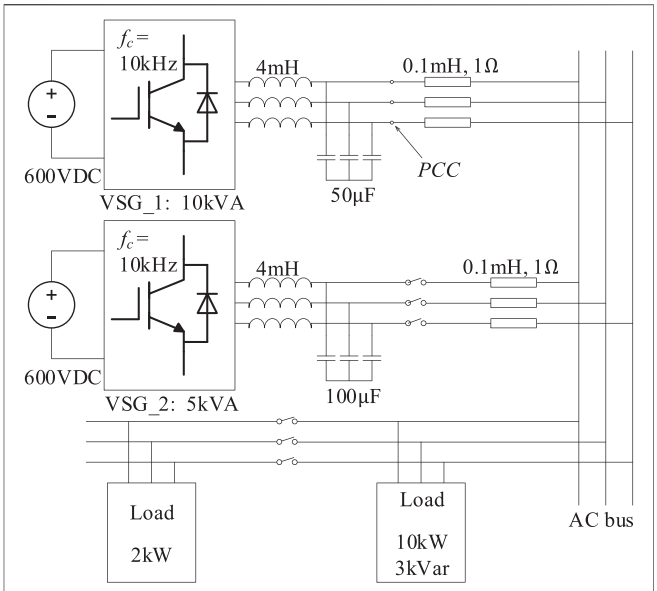

FIGURE 10 | Simulation topology of two VSGs sharing power in islanding.

TABLE 2 | Parameters of the 5 kVA inverter.

Parameter	Value	Comments
S_n and S_{base}	5 kVA	
V_n and V_{base}	220 V	
$\omega_n \approx \omega_0$	100 π rad/s	f = 50 Hz
X	0.63 Ω	L = 4 mH, $K = 3V_n^2/X$ = 115,546
k_p	318.5	$k_p^* = k_p\omega_n/S_{base}$, k_p^* = 20 p.u.
J	0.2	$M^* = J\omega_0^2/S_{base}$, M^* = 4 s
D	3,492	ζ = 1/sqrt(2)
k_p'	3,810.5	$k'_p = k_p + D$
H	11	$D = H{\cdot}k_p$
K_D	302.5	From eigenvalue loci similar to **Figure 6**

SIMULATION RESULTS

Simulation I: Single Virtual Synchronous Generator to Infinite Grid

In the grid-connected mode, the steady-state frequency is determined by the power grid. Assuming it is an infinite grid, the steady-state frequency would never change. The simulation topology is shown in **Figure 8**, and the control parameters are shown in **Table 1**.

In Situation I, the VSG could affect the frequency of PCC only in the transient state when the power reference P_{set} changes. Thus, we set P_{set} at 5kW at first; after the system reaches a steady state, P_{set} is then increased to 8kW. Four different control methods are applied in the same situation to compare the active power oscillation and angular frequency fluctuation. The simulation results are shown in **Figure 9**.

In **Figure 9**, the ideal basic VSG control method is not the "conventional scheme" but the "conventional scheme with an ideal PLL." An ideal PLL can only be obtained in simulation but cannot be realized in practice. A nonideal PLL in practice usually has an adverse effect on dynamic performance of the system. So, the simulation of "conventional scheme with an ideal PLL" is obviously better than that of "conventional scheme" on dynamic performance.

In the proposed method, the ω_{un} and ζ are also set at the same values as the ideal basic VSG control to mimic its dynamic properties.

The dynamic response of the ideal basic VSG control method is much better than that of the conventional scheme. The former can be considered as the target of the VSG control method. From **Figure 9**, we notice that the dynamic response of the proposed method is very close to that of the ideal basic method, and some indicators are even better. Thus, we can conclude that the proposed method is better than the conventional scheme in this situation.

Simulation II: Multi-VSGs With the Same Control Method

We use two parallel VSGs to simulate the parallel multi-VSG system. The simulation topology is shown in **Figure 10**.

VSG_1 is a 10-kVA inverter, and its main parameters are shown in **Table 1**. VSG_2 is a 5-kVA inverter, and its main parameters are shown in **Table 2**. Although the actual values of their most corresponding parameters are different, the per-unit values are almost all the same.

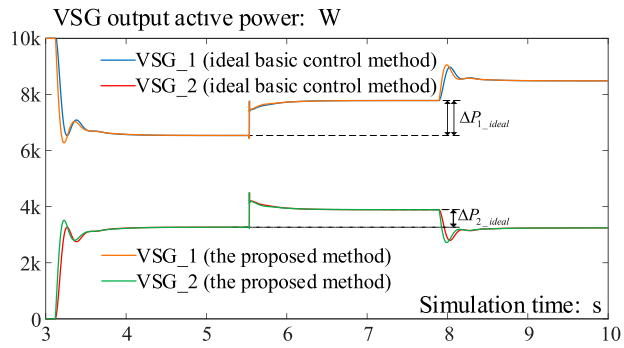

FIGURE 11 | Simulation results II: multi-VSGs with the same control method.

At first, VSG_1 is set to output 8 kW active power, which is 80% of its rated power. The load is set to consume 10 kW active power. On 3 s, VSG_2 prepares to connect to the grid, and connects to the grid after about 0.1 s; its output active power is set to 4 kW, which is also 80% of its rated power. Then, another 2 kW load is connected to the grid on about 5.5 s and the load steps to 12 kW in total. At last, the target power of VSG_1 is increased to 10 kW on about 8 s, and the target power of VSG_2 is maintained at 4 kW.

We did the simulation experiments twice. In the first simulation, both the VSGs are applied as the VSG ideal basic control method with an ideal PLL. In the second simulation, both the VSGs are applied as the proposed control method. In simulation, we can use an ideal PLL to get rid of the ill effect introduced through the actual PLL; however, the ideal PLL can never be realized in actual engineering, which makes the first simulation results much better than the real system. So, if the second simulation results are similar to those of the first one, it can certify that the proposed control method is better than the ideal basic control method, for the proposed one does not contain a PLL originally.

The simulation results are shown in **Figure 11**.

(1) The proposed control method has a similar dynamic response and steady operating point to the ideal basic control method.
(2) On about 5.5 s, the load is increased from 10 to 12 kW. At that very moment, no matter which control method is launched, each parallel VSG increases its output equally, which is related to the equal line impedance. Then, different control methods make differences to adjust the active power and achieve power-sharing. Finally, in the ideal basic control method and the proposed one, $\Delta P_1 \approx 2{\cdot}\Delta P_2$, which also means that the output power ratios to the rated power of the two VSGs are almost the same; thus, power-sharing is achieved well when the load is changed.
(3) On about 8 s, the load and the target power of VSG_2 remain unchanged, and the target power of VSG_1 turns to 10 kW. We can notice that the output angular frequencies of both VSGs increase and the VSG_1 increases. Because the angular frequency of VSG_2 is forced to increase by VSG_1, a few seconds later, the output power of VSG_1 does not reach 10 kW and the output power of VSG_2 decreases. This is the result of P-ω droop and power-sharing.

We can conclude from this simulation that the proposed method performs similarly to the ideal VSG control method. We can reasonably infer that the proposed method performs better than the conventional method in the parallel multi-VSG system.

Simulation III: The Cooperation Between the Proposed Method and the Ideal Basic One

The most important feature of VSG is that it can achieve power-sharing without using communication; thus, in this subsection, we use two parallel VSG-controlled inverters with diffident rated powers to demonstrate the power-sharing effect. VSG_1 (10 kV A) applies the ideal basic VSG control method, and VSG_2 (5 kV A) applies the proposed method. The topology of the simulation is shown in **Figure 12**. The control parameters of the two VSGs are listed in **Tables 1, 2**.

At first, VSG_1 is set to output 8kW active power, which is 80% of its rated power. The load is set to consume 10 kW active power. On 3 s, VSG_2 starts to prepare to connect to the grid and completes connecting after about 0.1 s; its output active power is set to 4 kW, which is also 80% of its rated power. Then another 2 kW load is connected to the grid on about 5.5 s, making the load step to 12 kW in total. At last, the power reference of VSG_1 steps up to 10 kW on about 8 s, and the power reference of VSG_2 remains at 4 kW. The simulation results are shown in **Figure 13**.

(1) On about 5.5 s, the load steps up. The proposed approach finally achieves power-sharing with the basic VSG control method because in **Figure 10**, $\Delta P_1 \approx 2{\cdot}\Delta P_2$ and $\Delta P_1/S_{base\_1} \approx \Delta P_2/S_{base\_2}$.
(2) Similar results can be seen after 8 s; when the power reference of VSG_1 steps up, the angular frequency

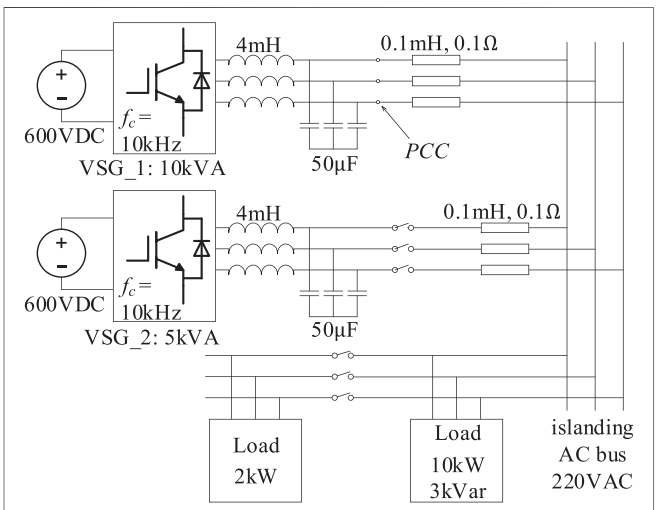

FIGURE 12 | Simulation topology of two VSGs sharing power in islanding.

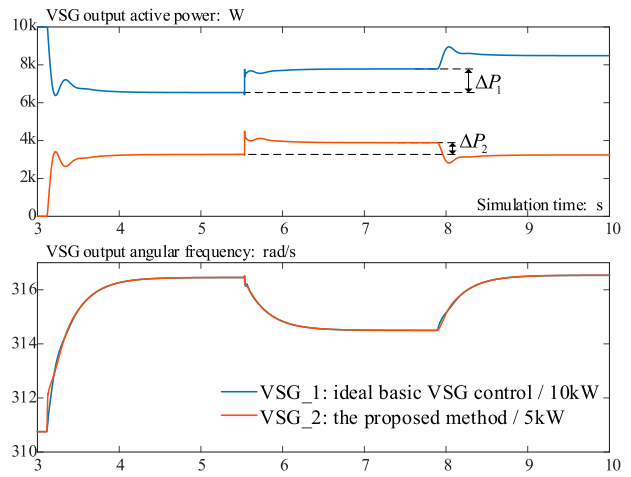

FIGURE 13 | Simulation results III: the proposed method and the ideal basic method.

changes a little, which results that VSG_2 shares some of the power increase.

Simulation III verifies that the proposed approach has a good power-sharing ability without using frequency detector.

In practice, the infinite grid is rare; most VSGs are connected to a non-infinite grid formed mainly by RSGs. Since the ideal basic VSG control method is the imitation of RSG, this simulation can also be treated as a VSG connected to an RSG or practical non-infinite grid.

CONCLUSION

This study aimed to investigate the parallel operation of VSGs with a damping correction term added to the swing equation. The addition of the extra term intends to achieve better oscillation damping in a multi-VSG environment. The load changing transients and eigenvalue analysis were employed to validate the transient and small-signal stability of the system with the proposed strategy.

In this article, a novel VSG control method is proposed for frequency fluctuation attenuation. The proposed method does not take grid angular frequency into computation, so that it can get rid of the ill effect introduced through PLL, which makes it especially fit for the parallel multi-VSG system. The damping power item is reconstrued with a new means to judge whether the system is in the steady state or not. The parametric design method is also figured out. At last, the simulation results to verify that the proposed method performs better than the conventional one.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusion of this article will be made available by the authors, without undue reservation.

AUTHOR CONTRIBUTIONS

ZS has done the main theory research work. FZ has done some of the simulation work. XC has done the work of article writing.

REFERENCES

Alipoor, J., Miura, Y., and Ise, T. (2015). Power System Stabilization Using Virtual Synchronous Generator with Alternating Moment of Inertia. *IEEE J. Emerg. Sel. Top. Power Electron.* 3 (2), 451–458. doi:10.1109/jestpe.2014.2362530

Alipoor, J., Miura, Y., and Ise, T. (2018). Stability Assessment and Optimization Methods for Microgrid with Multiple VSG Units. *IEEE Trans. Smart Grid* 9 (2), 1462–1471. doi:10.1109/tsg.2016.2592508

Alsiraji, H. A., and El-Shatshat, R. (2017). "Comprehensive Assessment of Virtual Synchronous Machine-Based Voltage Source Converter Controllers", *IET Generation. Transm. Distribution*, 11, 1762–1769.

Aouini, R., Marinescu, B., Ben Kilani, K., and Elleuch, M. (2016). Synchronverter-Based Emulation and Control of HVDC Transmission. *IEEE Trans. Power Syst.* 31 (1), 278–286. doi:10.1109/tpwrs.2015.2389822

D'Arco, S., and Suul, J. A. (2014). Equivalence of Virtual Synchronous Machines and Frequency-Droops for Converter-Based MicroGrids. *IEEE Trans. Smart Grid* 5 (1), 394–395. doi:10.1109/tsg.2013.2288000

Driesen, J., and Visscher, K. (2008). Virtual Synchronous Generators. in Proc. IEEE Power Energy Soc. Gen. Meeting-Conversion and Delivery of Electrical Energy in the 21st Century, 1–3.

Hafner, M., Finken, T., Felden, M., and Hameyer, K. (2011). Automated Virtual Prototyping of Permanent Magnet Synchronous Machines for HEVs. *IEEE Trans. Magn.* 47 (5), 1018–1021. doi:10.1109/tmag.2010.2091675

Hwang, M., Muljadi, E., Jang, G., and Kang, Y. C. (2017). Disturbance-adaptive Short-Term Frequency Support of a DFIG Associated with the Variable Gain Based on the ROCOF and Rotor Speed. *IEEE Trans. Power Syst.* 32 (3), 1873–1881. doi:10.1109/tpwrs.2016.2592535

Li, D., Zhu, Q., Lin, S., and Bian, X. Y. (2017). A Self-Adaptive Inertia and Damping Combination Control of VSG to Support Frequency Stability. *IEEE Trans. Energ. Convers.* 32 (1), 397–398. doi:10.1109/tec.2016.2623982

Li, D., Zhu, Q. S., and Bian, X. Y. (2017). A Self-Adaptive Inertia and Damping Combination Control of VSG to Support Frequency Stability. *IEEE Trans. Energ. Convers.* 32 (1), 397–398. doi:10.1109/tec.2016.2623982Li

Ma, Y., Cao, W., Yang, L., Wang, F. F., and Tolbert, L. M. (2017). Virtual Synchronous Generator Control of Full Converter Wind Turbines with Short-Term Energy Storage. *IEEE Trans. Ind. Electron.* 64 (11), 8821–8831. doi:10.1109/tie.2017.2694347

Mo, O., D'Arco, S., and Suul, J. A. (2017). Evaluation of Virtual Synchronous Machines with Dynamic or Quasi-Stationary Machine Models. *IEEE Trans. Ind. Electron.* 64 (7), 5952–5962. doi:10.1109/tie.2016.2638810

Shi, K., Ye, H., Song, W., and Zhou, G. (2018). Virtual Inertia Control Strategy in Microgrid Based on Virtual Synchronous Generator Technology. *IEEE Access* 6, 27949–27957. doi:10.1109/access.2018.2839737

Wang, F., Zhang, L., Feng, X., and Guo, H. (2018). An Adaptive Control Strategy for Virtual Synchronous Generator. *IEEE Trans. Ind. Applicat.* 54 (5), 5124–5133. doi:10.1109/tia.2018.2859384

Wang, F., Zhang, L., Feng, X., and Guo, H. (2018). An Adaptive Control Strategy for Virtual Synchronous Generator. *IEEE Trans. Ind. Appl.* 54 (5), 5124–5133.

Wang, S., Jing, L., Zhao, Y., Wickramasinghe, H. R., Wu, X., and Konstantinou, G. (2020). Operation of Unified Power Flow Controller as Virtual Synchronous Generator. *IEEE Access* 8, 162569–162580. doi:10.1109/access.2020.3021388

Wu, H., Ruan, X., Yang, D., Chen, X., Zhao, W., Lv, Z., et al. (2016). Small-signal Modeling and Parameters Design for Virtual Synchronous Generators. *IEEE Trans. Ind. Electron.* 63 (7), 4292–4303. doi:10.1109/tie.2016.2543181

Wu, W., Chen, Y., Luo, A., Zhou, L., Zhou, X., Yang, L., et al. (2017). A Virtual Inertia Control Strategy for DC Microgrids Analogized with Virtual Synchronous Machines. *IEEE Trans. Ind. Electron.* 64 (7), 6005–6016. doi:10.1109/tie.2016.2645898

Zhong, Q.-C., Nguyen, P., Ma, Z., and Sheng, W. (2014). Self-synchronized Synchronverters: Inverters without a Dedicated Synchronization Unit. *IEEE Trans. Power Electron.* 29 (2), 617–630. doi:10.1109/tpel.2013.2294425

Zhong, Q.-C., and Weiss, G. (2011). Synchronverters: Inverters that Mimic Synchronous Generators. *IEEE Trans. Ind. Electron.* 58 (4), 1259–1267. doi:10.1109/tie.2010.2048839

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Energy Research

ORIGINAL RESEARCH
published: 28 June 2021
doi: 10.3389/fenrg.2021.708131

Classification of Power Quality Disturbance Based on S-Transform and Convolution Neural Network

Jinsong Li[1]*, *Hao Liu*[1], *Dengke Wang*[2] and *Tianshu Bi*[1]

[1]North China Electric Power University, Beijing, China, [2]College of Information Science and Engineering, Northeastern University, Shenyang, China

The accurate classification of power quality disturbance (PQD) signals is of great significance for the establishment of a real-time monitoring system of modern power grids, ensuring the safe and stable operation of the power system and ensuring the electricity safety of users. Traditional power quality disturbance signal classification methods are susceptible to noise interference, feature selection, etc. In order to further improve the accuracy of power quality disturbance signal classification methods, this paper proposes a power quality disturbance classification method based on S-transform and Convolutional Neural Network (CNN). Firstly, S-transform is used to extract disturbance signals to obtain the time-frequency matrix with characteristics of the disturbance signals. As an extension of wavelet transform and Fourier transform, S-transform can avoid the disadvantages of difficult window function selection and fixed window width. At the same time, the feature extracted by S-transform has better noise immunity. Secondly, CNN is used to perform secondary feature extraction on the obtained high-dimensional time-frequency modulus matrix to reduce data dimensions and obtain the main features of the disturbance signal, then the main features extracted are classified by using the SoftMax classifier. Finally, after a series of simulation experiments, the results show that the proposed algorithm can accurately classify single disturbance signals with different signal-to-noise ratios and composite disturbance signals composed of single disturbance signals, and it also has good noise immunity. Compared with other classification methods, the algorithm proposed in this paper has better timeliness and higher accuracy, and it is an efficient and feasible power quality disturbance signal classification method.

Keywords: power quality disturbance, s-transform, convolutional neural network, feature extraction, noise immunity

OPEN ACCESS

Edited by:
Lei Xi,
China Three Gorges University, China

Reviewed by:
Wei Hu,
Zhejiang University, China
Rui Wang,
Nanyang Technological University, Singapore

***Correspondence:**
Jinsong Li
ljs050403@126.com

Specialty section:
This article was submitted to Smart Grids, a section of the journal Frontiers in Energy Research

Received: *11 May 2021*
Accepted: *07 June 2021*
Published: *28 June 2021*

Citation:
Li J, Liu H, Wang D and Bi T (2021) Classification of Power Quality Disturbance Based on S-Transform and Convolution Neural Network. Front. Energy Res. 9:708131. doi: 10.3389/fenrg.2021.708131

INTRODUCTION

In modern power systems, the rapid development of renewable energy power generation (Huang et al., 2021; Wang et al., 2021) and related distributed generations and microgrid control strategies (Huang et al., 2019; Wang et al., 2019) have injected a large number of nonlinear signals into the power system. At the same time, there are also a large number of nonlinear loads in the power grid (such as automotive charging piles, power transfer switches). The power grid is showing a power electronic trend, and the power quality problem of the distribution network is becoming more and more serious (Qiu et al., 2020). Frequent occurrences of power quality events cause a lot of economic losses and bring great inconvenience to people's lives. In order to deal with sudden power quality events, it is necessary to

accurately identify and classify the power quality disturbance signals. A convenient, fast and accurate classification algorithm can provide a higher-level application for modern smart meters and real-time monitoring system of power grid (Luo et al., 2018).

Current disturbance signal classification methods mainly include two steps:

1) Extracting characteristics of power quality disturbance signals;
2) Classifying with extracted features.

Feature extraction methods mainly include: Fast Fourier Transform (FFT) (Deng et al., 2020), Wavelet transform (Thirumala et al., 2018), S-transform (Kumar et al., 2015), Hilbert Huang transform (HHT) (Sun et al., 2018), short time Fourier transform (STFT) (Dhoriyani and Kundu, 2020), singular value decomposition (SVD) (Wang et al., 2017), Kalman filter (KF) (Niu et al., 2019). For step 1): due to relatively fixed length and shape of time window, short-time Fourier transform cannot reflect the characteristics of high frequency and low frequency. Although wavelet transform can realize multi-scale focusing, the relationship between transform scale and frequency is fixed. Singular value decomposition and Kalman filter lack the frequency domain characteristics of the signal. S-transform is a reversible time-spectrum positioning technology combining wavelet transform and FFT. It uses an analysis window, the width of the window changes with frequency to provide frequency-related resolution (Kumar et al., 2015). The time-frequency characteristics extracted by S-transform have more significant time-frequency characteristics (Tang et al., 2020).

In comparison, S-transform has higher time resolution and frequency resolution, and is more suitable for analyzing nonlinear, non-stationary, and transient power quality disturbances (Wang et al., 2021a).

The existing classifiers mainly include: artificial neural network (Haddad et al., 2018), Support Vector Machine (SVM) (Yong et al., 2015), decision tree (Huang et al., 2015; Long et al., 2018), expert system (Sai et al., 2015) and Bayesian classifier (Zhou et al., 2011), etc. For step 2): SVM has a high classification accuracy, but the amount of calculation in the process of parameter optimization is relatively large, and the real-time performance is not good. The expert system is a more flexible classification method, but with the increasing of different types of disturbance signals, the complexity of the knowledge base is getting higher and higher, which largely affects the fault tolerance of the system, and the classification performance is also restricted. In view of the problems of existing classifiers, finding a fast and accurate classification method has become the research focus of many researchers.

As the Frontier content in the field of artificial intelligence, neural networks have also made some preliminary applications in the field of power systems, and have achieved some remarkable results. In the field of electricity price forecasting, the literature (Jahangir et al., 2020) has greatly reduced the forecast error. Literature (Jiang et al., 2019) provides an intelligent fault diagnosis method that can automatically identify different health conditions of wind turbine gearboxes. Convolutional neural network (Convolution Neural Network, CNN), as a deep learning method of supervised learning, has advantages of low model complexity and fast calculation speed. Its unique convolution structure can reduce the amount of memory occupied by the deep network and the number of network parameters. CNN has been widely used in face recognition, text recognition and target tracking, as well as semantic segmentation and other fields (Chang et al., 2016; Chowdhury et al., 2016; Chen et al., 2018). In addition, CNN has excellent overfitting treatment methods compared to other classification methods. Methods such as reducing the number of network layers, using Dropout, and adding regular items can be used to improve overfitting.

However, in the field of power quality disturbance classification, the application of CNN is still immature. Only a small amount of literatures use CNN to solve the problem of power quality disturbance signal classification (Chen et al., 2018; Hezuo et al., 2018; Zhu et al., 2019). For example, literature (Chen et al., 2018) uses phase space reconstruction to reconstruct one-bit time series into a multi-dimensional space, then further project the obtained disturbance signal to a two-dimensional phase plane to form a two-dimensional trajectory image, finally input the trajectory image to a CNN for classification. Literature (Hezuo et al., 2018) maps the feature signal into a two-dimensional grayscale image, and then inputs it into a CNN for classification. Literature (Zhu et al., 2019) uses encoding and decoding to extract features of power quality disturbance signals, and then inputs the extracted features into a CNN for classification. However, it is difficult to distinguish the disturbance signal features with high similarity (such as interruption and sag) in the existing methods, and the signal feature extraction process also extracts many features which are irrelevant to disturbance signals. Although the existing methods have high classification accuracy, they still have certain misclassification phenomena.

In view of the above problems, this paper uses the combination of S-transform and CNN to classify power quality disturbance signals. The S-transform is used to extract the characteristic matrix which is used to represent the power quality disturbance signal. According to the three-dimensional (3D) network diagram of each disturbance signal, the sampling range of the feature vector corresponding to the disturbance signal that best represents the disturbance signal is determined. The matrix is trimmed to eliminate the eigenvectors that are useless for specific disturbance signal identification, that is, irrelevant vectors. And then get a square matrix that can represent the characteristics of the disturbance signal and the dimension is 125×125. Input the obtained square matrix into the CNN, and use the CNN to classify the power quality disturbance signal. The combination of S-transform and CNN to classify power quality disturbance signals ensures the efficiency, accuracy and robustness of the classification, and at the same time reduces the misclassification of disturbance signals, which is useful for establishing a real-time monitoring system for modern power grids. It is of great significance to ensure the safe and stable operation of the power system and ensure the safety of users' electricity.

S-TRANSFORM AND FEATURE EXTRACTION

The S-transform proposed by Stockwell (Stockwell et al., 1996) can be regarded as an extension of short-time Fourier transform and wavelet transform, and it is a reversible time-frequency analysis method.

S-transform is one of the best techniques for signal processing of non-stationary signals. It uses the phase information of continuous wavelet transform to correct the phase of the original wavelet. It can perform multi-resolution analysis on the signal, just like a set of filters with constant bandwidths. It uniquely has the frequency-related resolution, while positioning the real and imaginary spectra of the phase spectrogram. The time-frequency localization characteristics provided by S-transform are used for subsequent calculations.

Use the FFT and convolution theorem to calculate the S-matrix for each power quality disturbance time. The output of the S-matrix is a complex matrix whose dimension is $k \times n$, and the matrix expression is as follows

$$S(\tau,f) = A(\tau,f)e^{-i\varphi(\tau,f)} \tag{1}$$

where $A(\tau,f)$ represents amplitude, $\varphi(\tau,f)$ represents the phase.

The rows of the S matrix represent frequency, and the columns represent time. Each column represents the frequency component that appears in the signal at a specific time, and each row represents a specific frequency signal that occurs at the time from 0 to N–1 on each sampling point. The specific calculation method of S-transform is as follows.

Continuous S-Transform

The continuous S-transformation of the signal $h(t)$ is

$$S(t,f) = \int_{-\infty}^{+\infty} h(t)w(t-\tau,f)e^{-j2\pi f\tau}d\tau \tag{2}$$

where w is the Gaussian window function, expressed as

$$w(t-\tau) = \frac{|f|}{\sqrt{2\pi}}e^{\frac{-(t-\tau)^2f^2}{2}} \tag{3}$$

Discrete S-Transform

The power quality disturbance signal $h(t)$ can be discretized as $h(kT)$, T is the sampling interval; the Fourier transform form of the discrete sampling signal is

$$H\left[\frac{n}{NT}\right] = \frac{1}{N}\sum_{k=0}^{N-1} h(kT)e^{\frac{-i2\pi nk}{N}} \tag{4}$$

where $n = 0, 1, \cdots N-1$.

Let $\tau \to jt, f \to \frac{n}{(NT)}$, the improved discrete S-transform expression is as follows

$$\begin{cases} S\left[jT,\dfrac{n}{NT}\right] = \displaystyle\sum_{m=0}^{N-1} H\left[\frac{m+n}{NT}\right]G(m,n)e^{\frac{i2\pi mk}{N}}, n \neq 0 \\ S[jT,0] = \dfrac{1}{N}\displaystyle\sum_{m=0}^{N-1} h\left(\frac{m}{NT}\right), n = 0 \end{cases} \tag{5}$$

where $j, m, n = 0, 1, \ldots N-1, G(m,n) = e^{-\left(\frac{2n^2m^2}{n^2}\right)}$.

Time-Frequency Matrix Extraction and Cropping

It can be seen from the above that for a given power quality disturbance signal sequence, using S-transform to perform feature extraction on the sequence, a 2D matrix can be extracted, the row information of which represents the frequency feature and the column information for the time feature. Then, a 3D mesh graph of disturbance signal is made according to the extracted 2D matrix.

The dimension setting of the characteristic matrix is based on certain rules: after feature extraction of the source signal, a large number of feature vectors will be obtained, most of which are redundant features. Feature redundancy causes too many dimensions, will increase the amount of calculation, cause overlap of the features and misclassification. If the dimensionality is too few through dimensionality reduction, characteristics of the disturbance signal will be insignificant and the classification accuracy will decrease. Therefore, choosing an appropriate time-frequency matrix dimension is very important for the subsequent classification accuracy. Based on the CNN model of the TensorFlow platform, when reading the feature matrix, each feature matrix needs to be integrated into a line of a csv file. The maximum number of columns that the csv file can display is 16,384, and extra data cannot be displayed. When the maximum number of columns exceeds 16,384, the data will lead to not insert labels. In summary, this matrix 126×251 is selected for the dimension of a single input, this dimension can display the characteristics of the time-frequency matrix well without increasing the computational complexity.

In order to facilitate the subsequent input of the feature matrix into the CNN, the extracted initial feature matrix needs to be trimmed. **Figures 1**–**8** is a 3D mesh graph of each power quality signal sequence made by S-transform. In the figure, the x-axis coordinate is the number of sampling points, the y-axis is the frequency in Hz, and the z-axis is the normalized amplitude of the signal. Different colors indicate the degree of normalized amplitude, the lighter the color, the bigger the amplitude. Take the harmonic signal of **Figure 3** as an example, it is expressed as adding other harmonic components of different amplitudes on the basis of the normal signal. There are certain thresholds for the frequency and amplitude of the disturbance signal. By determining all types of disturbance signals within a certain range, the 3D mesh graph of each disturbance signal is compared with the 3D mesh graph of the normal signal, and finding the sampling range that best represents the characteristics of the disturbance signal. The feature matrix is trimmed according to the obtained sampling range. According to the obtained sampling range, the feature matrix is trimmed to obtain a square matrix of 125×125 as the input to the CNN. By trimming the feature matrix, the dimensionality of the input matrix and the interference can be reduced, and the classification accuracy and calculation speed can be improved.

CONVOLUTIONAL NEURAL NETWORK

Convolutional Neural Network (CNN), as a deep learning method, has been widely used in the field of pattern recognition and image classification. The weight sharing mechanism of CNN is very similar to the model of biological

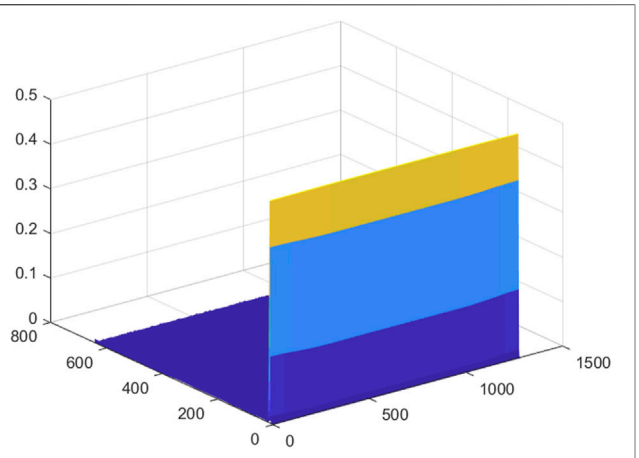

FIGURE 1 | S-transformation 3D mesh graph of normal signal.

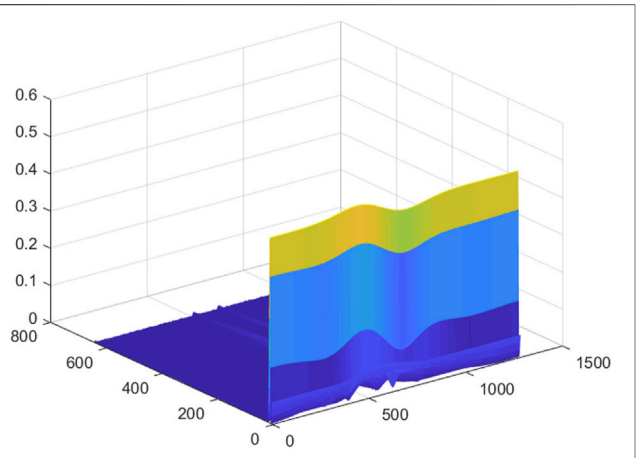

FIGURE 2 | S-transformation 3D mesh graph of transient pulse signal.

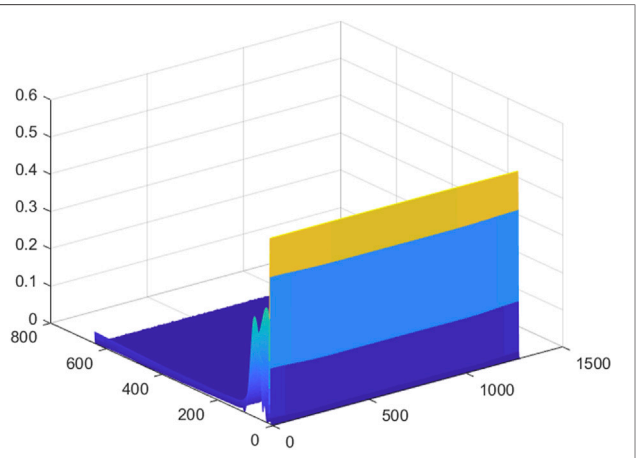

FIGURE 3 | S-transformed 3D mesh graph of harmonic signal.

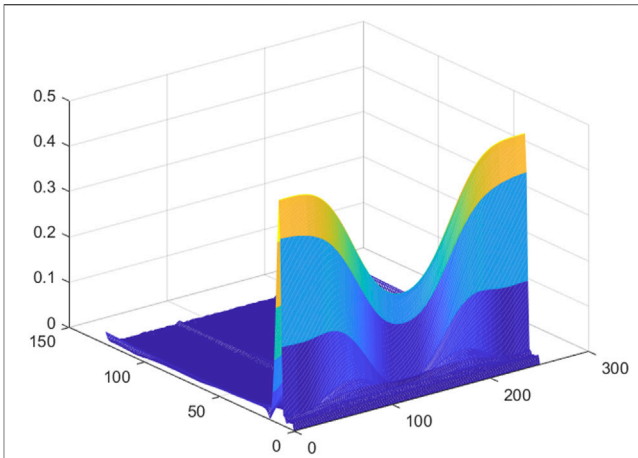

FIGURE 4 | S-transformation 3D mesh graph of the sag signal.

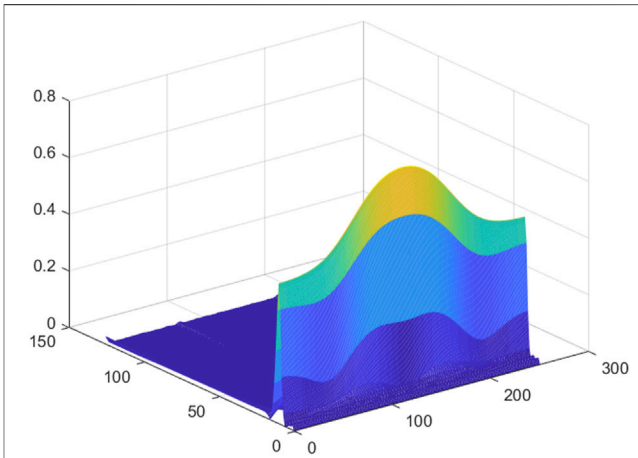

FIGURE 5 | S-transformed 3D mesh graph of the swell signal.

neural network. This mechanism makes the network model simpler and greatly reduces the number of weights (Chen et al., 2018). CNN is mainly composed of input layer, convolutional layer, pooling layer (down-sampling layer), and fully connected layer.

CNN Network Structure and Principle

The common CNN network is the LeNet-5 network, and its structure is shown in **Figure 9**. The first few stages need to extract features through multi-layer convolution.

The main components of CNN:

Convolutional layer: The purpose of the convolution operation is to extract different features of the input. The first convolutional layer may only extract some low-level features such as edges, lines, and corners. More layers of the network can iterate from the low-level features Extract more complex features.

Pooling layer: It is a form of downsampling. There are many different forms of non-linear pooling functions, of which Max-

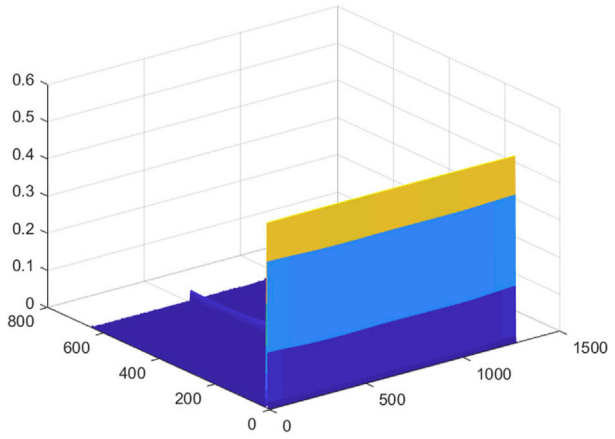

FIGURE 6 | S-transformation 3D mesh graph of transient oscillation signal.

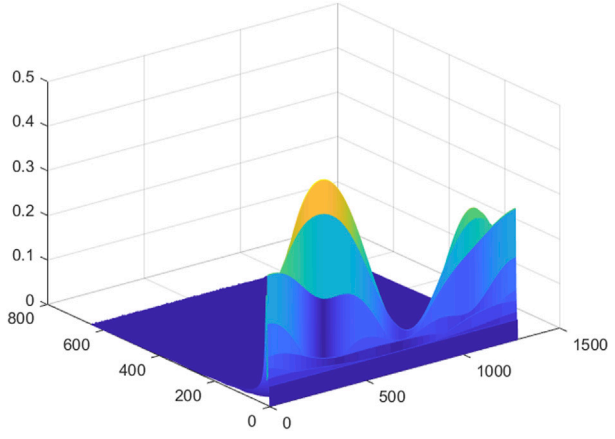

FIGURE 7 | S-transformed 3D mesh graph of the interrupt signal.

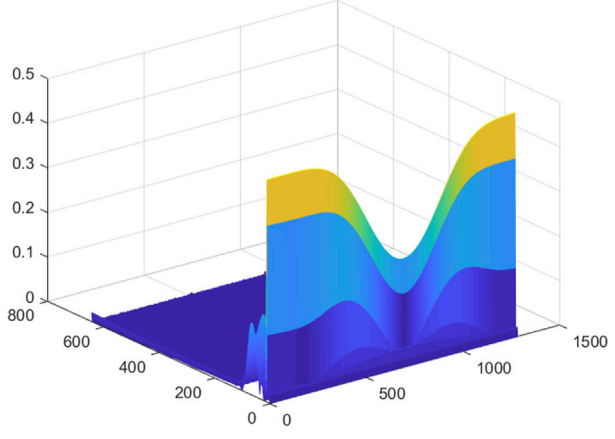

FIGURE 8 | S-transformation 3D mesh graph of sag and harmonic signal.

pooling and average sampling are the most common; the Pooling layer is equivalent to converting a higher resolution picture into a lower resolution picture; the pooling layer can Further reduce the number of nodes in the final fully connected layer, so as to achieve the purpose of reducing the parameters in the entire neural network.

Fully connected layer: The connection method is the same as that of a normal neural network, usually in the last few layers.

Generally speaking, CNN is a hierarchical model whose input is raw data, such as RGB images, raw audio data, etc. CNN extracts high-level semantic information from the original data through convolution, pooling, and nonlinear activation function mapping, and abstracts the original data layer by layer.

Convert the input raw data into the data form of a two-dimensional matrix, input it to the convolutional layer through the input layer, and use the convolutional layer to convolve the two-dimensional matrix. The calculation formula is as follows

$$y_i^m = g\left(\sum_{j=1}^{k} conv2D\left(y_i^{m-1}, \omega_{ij}^{m-1}\right) + b_i^m\right) \tag{6}$$

where $g()$ is the activation function, b_i is the bias value, ω_{ij} is the weight between neurons, and y_i is the ith input of the neuron.

Due to the slow convergence speed of the saturated nonlinear function, and even the problem of the disappearance of the gradient in the back propagation stage, the excitation function in this paper adopts the ReLu nonlinear function, and its expression is as follows

$$g(x) = \max(0, x) \tag{7}$$

After the original two-dimensional matrix is convolved by the convolution layer, the two-dimensional matrix obtained by the convolution operation is calculated by the ReLu activation function, and the calculated result is input to the pooling layer, and the downsampling operation is performed. As shown in the formula

$$y_i^m = down\left(y_i^{m-1}\right) + b_i^m \tag{8}$$

where *down*() represents the downsampling function.

By merging and pooling, the dimensionality of the input feature matrix is reduced, and the calculation amount of the network model is reduced. The fully connected layer is used to transfer the weights and biases between neurons in each layer, and finally is classified by the SoftMax classification layer.

Network Training Process

The CNN training process consists of two stages: the forward propagation stage (Forward) and the backward propagation stage (Backward).

Forward propagation stage: The input signal is continuously processed by convolution, pooling and activation function in the forward propagation stage, and the output O of the network is calculated layer by layer. Network calculation can be expressed as

$$O = G_n(\cdots(G_2(G_1(XW_1))W_2)\cdots W_n) \tag{9}$$

where G_i represents the nonlinear transformation; and $W_i\,(i = 1, 2, \cdots, n)$ represents the weight of each weight layer.

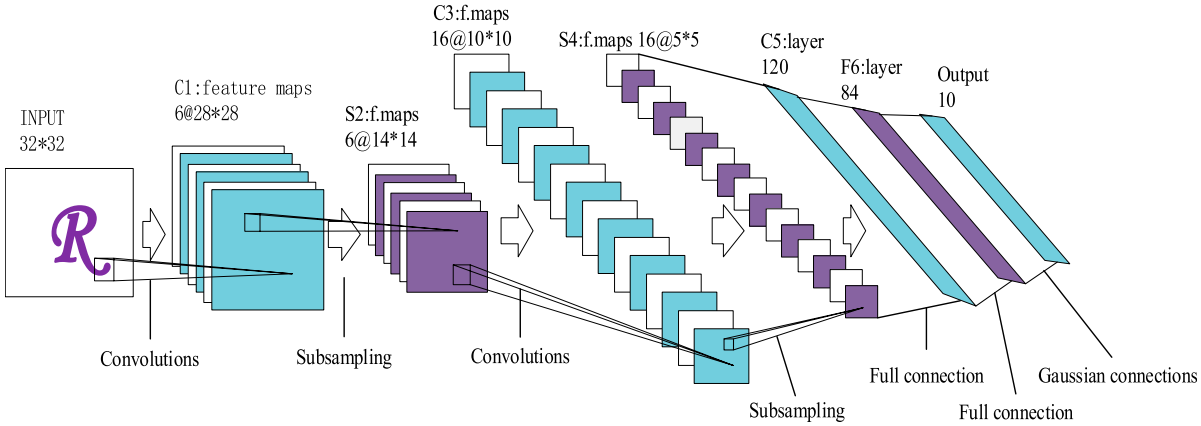

FIGURE 9 | LeNet-5 structure chart.

After getting the network output O, use the ideal output Y to evaluate the CNN network, and the ideal network satisfies $Y=O$.

Back propagation stage: According to the network output obtained in the forward propagation stage, the error is calculated, and the expression is as follows

$$E = \frac{1}{2}\|O - Y\|_2^2 \tag{10}$$

The gradient descent method is used to update and optimize the weights and bias coefficients between neurons in each layer of the network to minimize errors. The update method of weight and bias in the network model is shown in the following formula

$$W_{ij}^{l-1}(t+1) = W_{ij}^{l-1}(t) - \eta \frac{\partial E}{W_{ij}^{l-1}} \tag{11}$$

$$b_j^l(t+1) = b_j^l(t) - \eta \frac{\partial E}{\partial b_j^l} \tag{12}$$

where η represents the learning efficiency, E represents the error function.

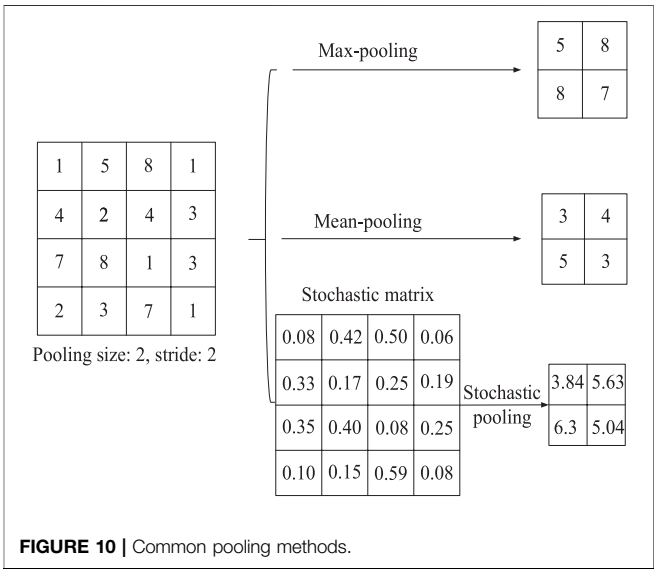

FIGURE 10 | Common pooling methods.

CNN Parameter Settings

For different classification tasks, the determination of the CNN structure requires both theoretical analysis and experimental observation to select appropriate parameters. Each network contains a different number of convolutional layers and corresponding pooling layers, and the parameter settings of each convolutional layer and pooling layer are also different.

The convolution kernel parameters that need to be set are: stride (sliding step size), padding (convolution method) and the size of the convolution kernel. Stride should not be set too large, because too large will result in the loss of the feature amount of the input data, so stride is generally set to 1 or 2. There are two modes of padding setting: *same* and *valid*, *same* means that after the convolution operation, the dimensionality of the input data remains unchanged (0-padding is performed on the periphery of the input data according to *stride*'s value); *valid* means that the dimensionality of the input data will be reduced correspondingly after the convolution operation, and the size of the convolution kernel is determined according to the dimensions of the input data. The calculation method of the output data size is as follows

$$U = \left[\frac{(I - C + 2*P)}{S}\right] + 1 \tag{13}$$

where U is the size of the output data, I is the size of the input data, C is the size of the convolution kernel, P is the number of zero padding, and S is the size of the stride.

The sole purpose of the pooling layer is to reduce the dimensionality of the input data, and its parameter settings are: the selection of the pooling method, the size of the pooling layer and the sliding step length. Take an example to introduce the size and sliding step length of the pooling layer: input a 4×4 data, set the size of the pooling layer to 2×2, and set the step length to 2, and get an output 2×2 data after pooling. **Figure 10** shows several common pooling methods.

Max-pooling only retains the maximum value in the area. Mean-pooling preserves the average value of the feature points in

the area. Stochastic pooling only needs to randomly select the elements in the feature map according to their probability value, and the probability of element selection is positively related to its value. Among them, Max-pooling retains the maximum value, ignoring other values, which can reduce the impact of noise, improve the robustness of the model, reduce the number of model parameters, help reduce model overfitting problems, and be more suitable for power quality classification problems.

EXAMPLE CONSTRUCTION

Mathematical Model of Power Quality Disturbance

The validity of real-time power quality disturbance data is affected by some other factors. For example, obtaining real-time power quality disturbance data requires a long monitoring time, and the location of the power quality disturbance event is uncertain, which greatly affects work efficiency. Therefore, using MATLAB to simulate the mathematical model of the power quality disturbance signal, the disturbance signal obtained by the simulation can accurately describe the real-time data in accordance with international standards (Chowdhury et al., 2016). Voltage sags, swells, spikes, interruptions, flickers, transient oscillations, harmonics, sags and harmonics, swells and harmonics are several common power quality disturbance signals. Attached schedule 1 is the model of 10 kinds of disturbance signals and standard signals, which are expressed as $S0, S1, \cdots S9$. among them $f = 50Hz$; $\omega = 2\pi f$; $T = \frac{1}{f}$.

Construction of Simulation Experiment Platform

This paper uses a two-dimensional CNN structure based on deep learning, uses TensorFlow deep learning framework, and Python 3.5 programming language to build a network model. The TensorFlow deep learning framework was built using a laptop equipped with a 64-bit Ubuntu Linux 16.04LTS system and NVIDIA GTX1080 graphics card. TensorFlow is an open-source software library that uses data flow graphs for numerical calculations. Its workflow is relatively easy, its API is stable, its compatibility is good, and it can be perfectly combined with NumPy. TensorFlow's compilation time is very short, it can be iterated faster, and its flexibility and efficiency are relatively high. Using TensorFlow to build a two-dimensional convolutional neural network model, the program compilation is simple, the simulation speed is relatively fast, the flexibility is high, and it can be well adapted to the numerical optimization task.

The CNN Model Used in This Article

The CNN model used in this paper is improved based on the traditional LeNet-5 architecture model, including two convolutional layers and two pooling layers. The parameter settings of two convolution kernels are different, the specific parameter settings of the first convolution kernel: stride is set to 1, padding is set to same, the size of the convolution kernel is 3×3. The parameter settings of the second convolution kernel: stride is set to 1, padding is set to same, and the size of the convolution kernel is 5×5. The parameter settings of the two pooling layers are the same. The specific parameter settings are: Max-pooling is selected as the pooling method, the size of the pooling layer is 5×5, and the step size is set to 5. The dimension of the data input in this paper is 125×125, after the convolution and pooling operation, the dimension of the output data obtained is 5×5, and the output data obtained is input into the fully connected layer for normalization processing to avoid the impact of classification with large data values. **Figure 11** shows the convolutional neural node pair network model used.

The cross-entropy loss function is used as the loss function of the CNN, and the SoftMax classification layer is used for classification. **Figure 12** shows the system structure model of this article.

In the field of machine learning, if the model has too many parameters and the number of training samples is too little, it will lead to overfitting of the trained model. Overfitting often occurs in the training process of neural networks, the specific performance is: the model has a small loss function and high prediction accuracy on the training data, while on the test data, the loss function is relatively large and the prediction accuracy is low. In order to prevent the occurrence of overfitting, the CNN model used in this paper adds the Dropout function. In the process of forward propagation, the Dropout function allows a certain neuron to stop working with a certain probability, which can make the generalization ability of the neural network model stronger, so that it will not rely too much on some local features.

The role of the Dropout function:

1) Averaging effect: The Dropout removes neurons in different hidden layers is similar to training different networks, and the Dropout is equivalent to averaging multiple different neural networks.
2) Reduce the complex co-adaptation relationship between neurons: The update of weights no longer depends on the joint action of hidden nodes with fixed relationships, forcing the network to learn more robust features.
3) Dropout is similar to the role of gender in biological evolution: In order to survive, species tend to adapt to the new environment and can breed new species that adapt to the environment. This behavior is similar to training an applicable network model, which effectively prevents overfitting.

Disturbance Signal Classification Process

The flow diagram of the classification of power quality disturbance signals is shown in **Figure 13**.

The specific steps are as follows:

1) Preprocess the power quality disturbance signal generated by MATLAB, use S-transform to extract the time-frequency matrix representing the disturbance signal, and draw a 3D network diagram of the disturbance signal.
2) According to the time-frequency matrix extracted from the 3D network graph of the disturbance signal, a new matrix of dimension 125×125 is obtained, and the training set is formed to train the CNN.

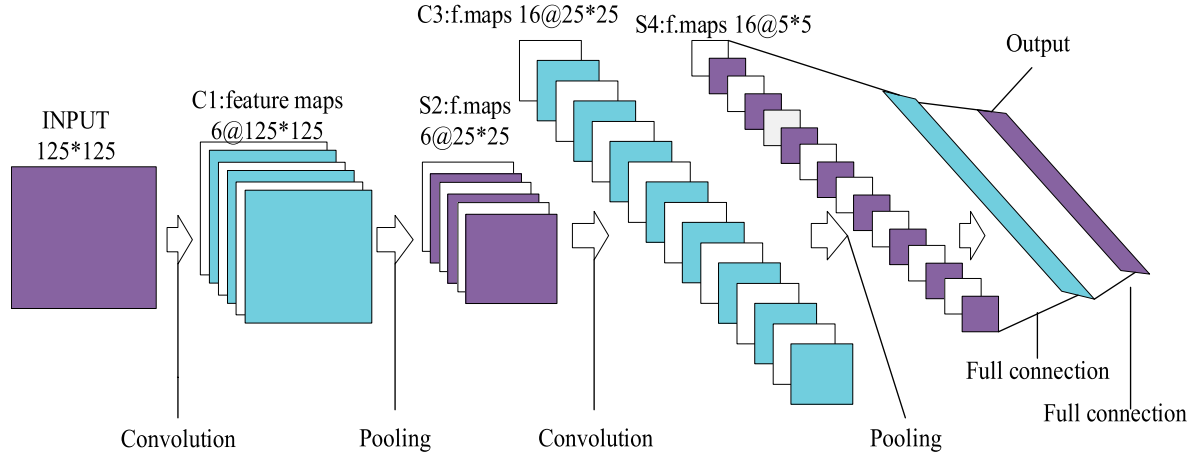

FIGURE 11 | CNN model.

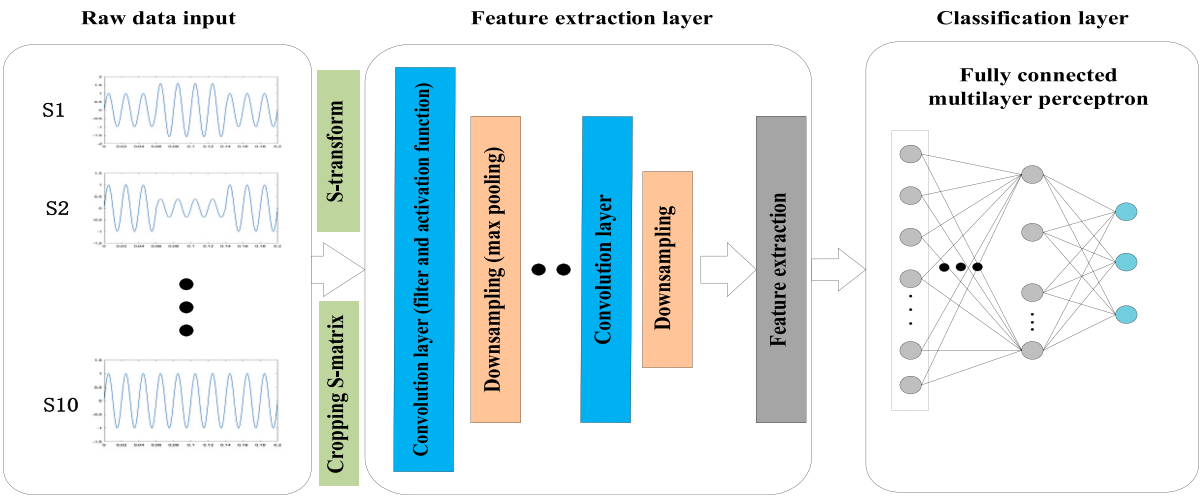

FIGURE 12 | System structure model presented in this paper.

3) The cross-entropy loss function is adopted, and the Dropout function is added in the forward propagation stage to prevent the occurrence of overfitting. Use stochastic gradient descent method to update the parameter model, and optimize the model through error back propagation.
4) After the input data is convolved and pooled, the characteristics of the disturbance signal are extracted, and the SoftMax classification layer is used for classification. Then the verification and test sets are used for verification and testing to obtain the final classification results.

SIMULATION AND ANALYSIS

CNN Training

This article uses MATLAB to generate the power quality signals shown in **Supplementary Table S1**. Normal signals and every type of disturbance signal each generates 500 random samples, a total of 5,000 samples, each signal is added with a signal-to-noise ratio (SNR) of 20, 30 and 40dB Gaussian white noise. The feature matrix of all power quality signals is extracted from S-transform, and the feature matrix is trimmed using a 3D mesh graph. The trimmed feature matrix is integrated into a row of feature values by row, and a digital label is added to each row of data (0–9, respectively represent the labels of 10 disturbance signals). Shuffle all the data in rows and extract the first 3,000 rows of data from the disrupted data set to form the training set, the middle 1,000 rows of data form the verification set, and the last 1,000 rows of data form the test set. Use CNN to read the csv file containing the disturbance signal data.

In order to evaluate the training status and training effect of the network, the cross-entropy loss function and the classification accuracy rate are drawn with the number of iterations (each epoch represents training 50 times), namely the loss function

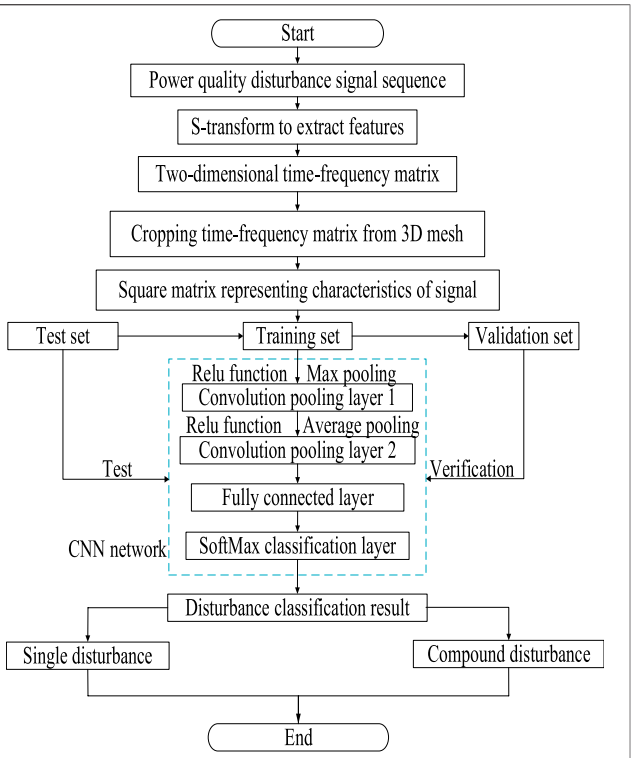

FIGURE 13 | Process diagram of classification of power quality disturbance signals.

TABLE 1 | Classification accuracy of CNN with different SNR.

Disturbance type	SNR/dB		
	20	30	40
Sinusoidal signal S0	1.000	1.000	1.000
Swell S1	0.985	0.988	0.995
Sag S2	0.987	0.990	0.991
Flicker S3	1.000	1.000	1.000
Transient pulse S4	0.986	0.989	0.993
Interrupt S5	0.984	0.987	0.989
Harmonics S6	0.983	0.989	0.994
Transient flicker S7	0.987	0.993	0.996
Swell and harmonics S8	0.991	0.993	0.997
Sag and harmonics S9	0.988	0.990	0.994
Average	0.989	0.992	0.995

curve and the classification accuracy curve. As shown in **Figure 14**, the loss function curve has a relatively large decline when the network is first trained. As the number of iterations increases, the loss function curve begins to fluctuate, but gradually stabilizes. As shown in **Figure 15**, the classification accuracy curve gradually increases as the number of iterations increases, and finally rises to a higher classification accuracy close to 1. As the number of iterations increases, the two curves gradually tend to converge, which proves that the entire network is continuously optimized and improved, and the stability of the network is gradually increasing. By comparing the classification effects of disturbance signals with different signal-to-noise ratios, it can be

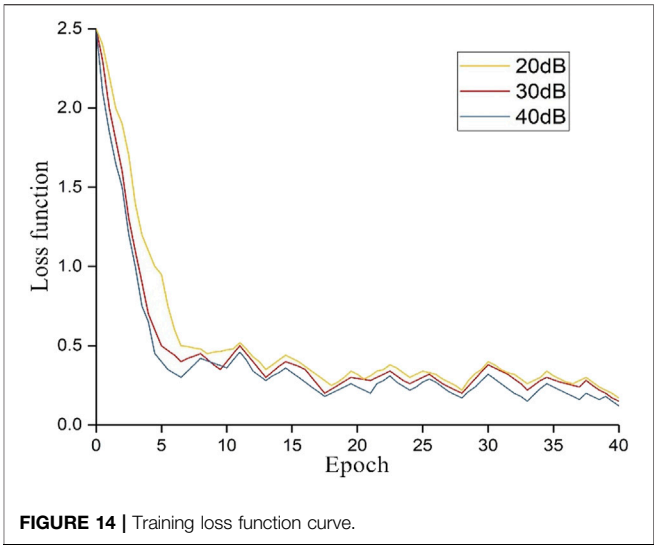

FIGURE 14 | Training loss function curve.

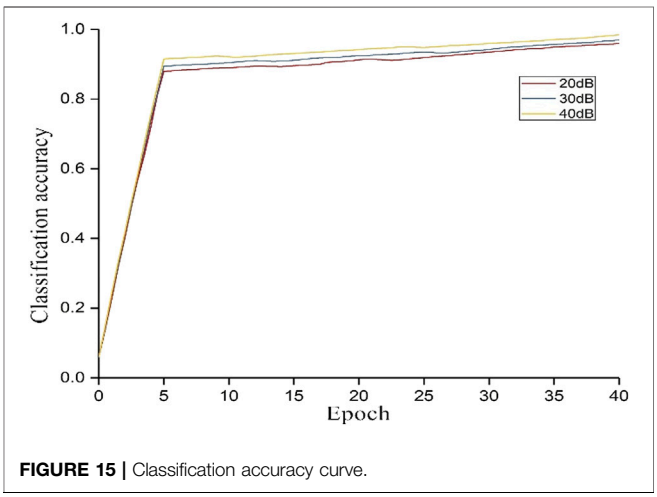

FIGURE 15 | Classification accuracy curve.

seen that the network still maintains a high classification accuracy rate for signals with different noises, indicating that the method has certain noise immunity and strong robustness.

Classification Effect

In order to further verify the effectiveness of this method, tests are performed under different noise intensities. The classification accuracy is shown in **Table 1**. It can be seen from **Table 1** that CNN has higher accuracy under different noise intensities, indicating that the proposed method has strong noise immunity performance in the classification of power quality disturbance signals. In order to further determine the misclassification of disturbance signals, take the case of a signal-to-noise ratio of 40dB as an example, and list the classification results of each disturbance signal in the table below. It can be found that the classification accuracy of each signal is relatively high, and there is no excessive misclassification. The specific classification results of various disturbance signals are shown in **Table 2**.

TABLE 2 | Classification result details when the SNR is 40dB.

Type	S0	S1	S2	S3	S4	S5	S6	S7	S8	S9
S0	100	0	0	0	0	0	0	0	0	0
S1	0	99	0	1	0	0	0	0	0	0
S2	0	1	98	0	0	1	0	0	0	0
S3	0	0	0	100	0	0	0	0	0	0
S4	0	0	0	0	99	0	1	0	0	0
S5	0	0	1	0	0	99	0	0	0	0
S6	0	0	0	0	0	0	99	1	0	0
S7	0	0	0	0	0	0	0	99	0	0
S8	0	1	0	0	0	0	0	0	99	0
S9	0	0	1	0	0	0	0	0	0	99

Comparative Analysis With Existing Classification Models

The proposed classification model and existing classification models are compared and analyzed to judge the classification effect of the classification model proposed in this paper. Models used for comparison include Probabilistic Neural Networks (PNN) (Zhengming et al., 2018), Principal Component Analysis-based Support Vector Machines (PCA-SVM) (Jiang et al., 2019a), and traditional Convolutional Neural Networks (CNN) (Song et al., 2018). The parameter setting of each model is set according to the existing reference documents, and will not be repeated here.

As shown in **Table 3**, it is the comparison result of the classification accuracy of different noise disturbance signals for each model. Comparing and analyzing the accuracy of different classification algorithms under different noise conditions, it is clear that the algorithm proposed in this article maintains a high classification accuracy rate under 20–40dB noise conditions. The results show that the classification accuracy of PNN and PCA-SVM is slightly lower than the model proposed in this paper. Since S-transform-CNN has an additional step of feature extraction using S transform, the model proposed in this paper has a higher classification accuracy and better noise immunity than traditional CNN model.

In addition to classification accuracy, this paper also compares classification time, the comparison results are shown in **Table 4**. It can be seen that the training time of PNN is relatively longer, because its structure is relatively complex and the number of neurons is relatively large, so the computational complexity is higher than the proposed method in this paper. The SVM in PCA-SVM belongs to binary classification, and the training and testing time is long. Since the proposed model has an extra feature extraction process compared with the traditional CNN, the training time is slightly longer.

From the comprehensive analysis results of the above two tables, it can be seen that when considering the two factors of accuracy and time consumption, the classification accuracy of the S-transform-CNN method proposed in this paper is slightly lower than that of PNN, but the time consumed is much less than that of PNN. The reason is that the number of neurons in the PNN is relatively large, which greatly increases the computational complexity and the time consumed by the network. Among the existing disturbance signal classification methods, most of the classification methods focus on off-line detection and disturbance classification of power quality disturbance signals. As power quality problems become more and more complex and users have higher and higher requirements for power quality, it is necessary to conduct online analysis of power quality problems, and a shorter classification time is even more important. Considering comprehensively, the method proposed in this paper has higher classification accuracy and lower Time-consuming, which indicates that it can reduce the time of network training and testing and improve work efficiency while ensuring the classification accuracy.

TABLE 3 | Classification accuracy of different algorithms.

Algorithm	Classification accuracy		
	20dB	30dB	40dB
PNN	0.986	0.989	0.992
PCA-SVM	0.965	0.968	0.971
CNN	0.952	0.954	0.958
S-transform-CNN	0.986	0.992	0.995

TABLE 4 | Time consumption comparisons of different algorithms.

Algorithm	Training time(s)	Testing time(s)	Total time(s)
PNN	637	1.1	638.1
PCA-SVM	468	1.3	469.3
CNN	193	0.6	193.6
S-transform-CNN	205	0.8	205.8

CONCLUSION

This paper proposes a new method of power quality disturbance classification based on S-transform and CNN. Use S-transform to extract characteristics of disturbance signals, extract the time-frequency matrix representing the characteristics of the disturbance signal, then use the 3D mesh graph of the disturbance signal to trim the extracted matrix, and input the processed matrix into the CNN for classification. Under different noise levels, this method obtains relatively good classification accuracy for power quality disturbance signals, and has good noise immunity. The difference between this method and other methods based on CNN is the input form of the CNN. Traditional methods input the gray image of the disturbance signal. This paper directly inputs the characteristic matrix of the disturbance signal into the CNN. Compared with the traditional method, the method in this paper is more concise and reduces the loss of characteristics. Under the premise of ensuring classification accuracy and noise immunity. Further research will try to improve the performance of this method by introducing new feature extraction rules, and consider introducing more complex disturbance signals for classification to meet actual power quality analysis needs.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

JL conceived the idea for the manuscript and wrote the manuscript with input from HL, DW, and TB. All authors have read and agreed to the published version of the manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fenrg.2021.708131/full#supplementary-material

REFERENCES

Chowdhury, A. R., Lin, T.-Y., Maji, S., and Learned-Miller, E., (2016). "One-to-many Face Recognition with Bilinear CNNs," in 2016 IEEE Winter Conference on Applications of Computer Vision (WACV), pp. 1–9. doi:10.1109/WACV.2016.7477593

Chang, L., Deng, X. M., and Zhou, M. Q., (2016). Convolutional Neural Networks in Image Understanding. *Acta Automatica Sinica* 42 (9), 1300–1312. doi:10.16383/j.aas.2016.c150800

Chen, W., He, J., and Pei, X. (2018). Classification for Power Quality Disturbance Based on Phase-Space Reconstruction and Convolution Neural Network. *Dianli Xitong Baohu Yu Kongzhi/Power Syst. Prot. Control.* 46 (14), 87–93. doi:10.7667/PSPC171080

Deng, H., Gao, Y., Chen, X., Zhang, Y., Wu, Q., and Zhao, H. (2020). "Harmonic Analysis of Power Grid Based on FFT Algorithm," in 2020 IEEE International Conference on Smart Cloud (SmartCloud), 161–164. doi:10.1109/SmartCloud49737.2020.00038

Dhoriyani, S. L., and Kundu, P. (2020). "Comparative Group THD Analysis of Power Quality Disturbances Using FFT and STFT," in 2020 IEEE First International Conference on Smart Technologies for Power (Energy and Control (STPEC)), 1–6. doi:10.1109/STPEC49749.2020.9297759

Haddad, R. J., Guha, B., Kalaani, Y., and El-Shahat, A. (2018). Smart Distributed Generation Systems Using Artificial Neural Network-Based Event Classification. *IEEE Power Energ. Technol. Syst. J.* 5 (2), 18–26. doi:10.1109/JPETS.2018.2805894

Hezuo, Q. U., Xiaoming, L. I., and Chen, C., (2018). *Classification of Power Quality Disturbances Using Convolutional Neural Network*. Engineering Journal of Wuhan University.

Huang, B., Li, Y., Zhan, F., Sun, Q., and Zhang, H. (2021). A Distributed Robust Economic Dispatch Strategy for Integrated Energy System Considering Cyber-Attacks. *IEEE Trans. Ind. Inf.*, 1. in press. doi:10.1109/TII.2021.3077509

Huang, B., Liu, L., Li, Y., and Zhang, H. (2019). Distributed Optimal Energy Management for Microgrids in the Presence of Time-Varying Communication Delays. *IEEE Access* 7, 83702–83712. doi:10.1109/ACCESS.2019.2924269

Huang, N., Zhang, W., and Cai, G., (2015). Power Quality Disturbances Classification with Improved Multiresolution Fast S-Transform. *Power Syst. Tech.* 39 (05), 1412–1418. doi:10.13335/j.1000-3673.pst.2015.05.036

Jahangir, H., Tayarani, H., Baghali, S., Ahmadian, A., Elkamel, A., Golkar, M. A., et al. (2020). A Novel Electricity Price Forecasting Approach Based on Dimension Reduction Strategy and Rough Artificial Neural Networks. *IEEE Trans. Ind. Inf.* 16 (4), 2369–2381. doi:10.1109/TII.2019.2933009

Jiang, G., He, H., Yan, J., and Xie, P. (2019). Multiscale Convolutional Neural Networks for Fault Diagnosis of Wind Turbine Gearbox. *IEEE Trans. Ind. Electron.* 66 (4), 3196–3207. doi:10.1109/TIE.2018.2844805

Jiang, J., Wen, Z., Zhao, M., Bie, Y., Li, C., Tan, M., et al. (2019a). Series Arc Detection and Complex Load Recognition Based on Principal Component Analysis and Support Vector Machine. *IEEE Access* 7, 47221–47229. doi:10.1109/ACCESS.2019.2905358

Kumar, R., Singh, B., Shahani, D. T., Chandra, A., and Al-Haddad, K. (2015). Recognition of Power-Quality Disturbances Using S-Transform-Based ANN Classifier and Rule-Based Decision Tree. *IEEE Trans. Ind. Applicat.* 51 (2), 1249–1258. doi:10.1109/TIA.2014.2356639

Long, J. (2018). Feature Extraction and Classification of UHF PD Signals Based on Improved S-Transform. *Gaodianya Jishu/High Voltage Eng.* 44 (11), 3649–3656. doi:10.13336/j.1003-6520.hve.20181031026

Luo, Y., Li, K., Li, Y., Cai, D., Zhao, C., and Meng, Q. (2018). Three-Layer Bayesian Network for Classification of Complex Power Quality Disturbances. *IEEE Trans. Ind. Inf.* 14 (9), 3997–4006. doi:10.1109/TII.2017.2785321

Niu, S., Wang, K., and Liang, Z. (2019). Synchronous Phasor Estimation Method for Power System Based on Modified Strong Tracking Unscented Kalman Filter. *Power Syst. Tech.* 43 (09), 3218–3225. doi:10.13335/j.1000-3673.pst.2018.2814

Qiu, W., Tang, Q., Liu, J., and Yao, W. (2020). An Automatic Identification Framework for Complex Power Quality Disturbances Based on Multifusion Convolutional Neural Network. *IEEE Trans. Ind. Inf.* 16 (5), 3233–3241. doi:10.1109/TII.2019.2920689

Sai, T. K., and Reddy, K. A. (2015). New Rules Generation from Measurement Data Using an Expert System in a Power Station. *IEEE Trans. Power Deliv.* 30 (1), 167–173. doi:10.1109/TPWRD.2014.2355595

Song, H., Dai, J., and Zhang, W., (2018). Partial Discharge Pattern Recognition Based on Deep Convolutional Neural Network under Complex Data Sources. *Gaodianya Jishu/High Voltage Eng.* 44 (11), 3625–3633. doi:10.13336/j.1003-6520.hve.20181031023

Stockwell, R. G., Mansinha, L., and Lowe, R. P. (1996). Localization of the Complex Spectrum: the S Transform. *IEEE Trans. Signal. Process.* 44 (4), 998–1001. doi:10.1109/78.492555

Sun, Y., Tang, X., and Sun, X., (2018). Research on Multi-type Energy Storage Coordination Control Strategy Based on MPC-HHT. *Proc. CSEE* 38 (09), 2580–2588. doi:10.13334/j.0258-8013.pcsee.171042

Tang, Q., Qiu, W., and Zhou, Y. (2020). Classification of Complex Power Quality Disturbances Using Optimized S-Transform and Kernel SVM. *IEEE Trans. Ind. Electron.* 67 (99), 9715–9723. doi:10.1109/TIE.2019.2952823, PP(

Thirumala, K., Prasad, M. S., Jain, T., and Umarikar, A. C. (2018). Tunable-Q Wavelet Transform and Dual Multiclass SVM for Online Automatic Detection of Power Quality Disturbances. *IEEE Trans. Smart Grid* 9 (4), 3018–3028. doi:10.1109/TSG.2016.2624313

Wang, F., Quan, X., and Ren, L., (2021). Review of Power Quality Disturbance Detection and Identification Methods. *Proc. CSEE* 1-17. in press. http://kns.cnki.net/kcms/detail/11.2107.TM.20201119.0900.002.html.

Wang, R., Sun, Q., Gui, Y., and Ma, D. (2019). Exponential-function-based Droop Control for Islanded Microgrids. *J. Mod. Power Syst. Clean. Energ.* 7 (4), 899–912. doi:10.1007/s40565-019-0544-3

Wang, R., Sun, Q., Tu, P., Xiao, J., Gui, Y., and Wang, P. (2021a). Reduced-Order Aggregate Model for Large-Scale Converters with Inhomogeneous Initial Conditions in DC Microgrids. *IEEE Trans. Energ. Convers.*, 1. in press. doi:10.1109/TEC.2021.3050434

Wang, Y., Li, Q., and Zhou, F. (2017). A Novel Algorithm for Transient Power Quality Disturbances Detection. *Proc. CSEE* 37 (24), 7121–7132. doi:10.13334/j.0258-8013.pcsee.162592

Yong, D. D., Bhowmik, S., and Magnago, F. (2015). An Effective Power Quality Classifier Using Wavelet Transform and Support Vector Machines. *Expert Syst. Appl.* 42 (15-16), 6075–6081. doi:10.1016/j.eswa.2015.04.002

Zhengming, L. I., Qian, L., and Jiabin, L. I. (2018). Type Recognition of Partial Discharge in Power Transformer Based on Statistical Characteristics and PNN. *Power Syst. Prot. Control.* 46 (13), 55–60. doi:10.7667/PSPC170962

Zhou, L., Guan, C., and Lu, W. (2011). Application of Multi-Label Classification Method to Catagorization of Multiple Power Quality Disturbances. *Zhongguo Dianji Gongcheng Xuebao/Proceedings Chin. Soc. Electr. Eng.* 31 (4), 45–50. doi:10.1631/jzus.C1000008

Zhu, Ruijin., Guo, Weilin., and Gong, Xuejiao. (2019). Power Quality Disturbance Classification Based on Self Encoder and Convolutional Neural Network. *J. Power Syst. automation* 31, 70–75. doi:10.19635/j.cnki.csu-epsa.000199

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Energy Research

ORIGINAL RESEARCH
published: 27 July 2021
doi: 10.3389/fenrg.2021.695057

A Flexible Ensemble Algorithm for Big Data Cleaning of PMUs

Long Shen, Xin He, Mingqun Liu, Risheng Qin, Cheng Guo, Xian Meng and Ruimin Duan*

Electric Power Research Institute of Yunnan Power Grid Company Ltd., Kunming, China

With an increasing application of Phase Measurement Units in the smart grid, it is becoming inevitable for PMUs to operate in severe conditions, which results in outliers and missing data. However, conventional techniques take excessive time to clean outliers and fill missing data due to lacking support from a big data platform. In this paper, a flexible ensemble algorithm is proposed to implement a precise and scalable data clean by the existing big data platform "Apache Spark." In the proposed scheme, an ensemble model based on a soft voting approach utilizes principal component analysis in conjunction with the K-means, Gaussian mixture model, and isolation forest technique to detect outliers. The proposed scheme uses a gradient boosting decision tree for each extracted feature of PMUs for the data filling process after detecting outliers. The test results demonstrate that the proposed model achieves high accuracy and recall by comparing simulated and real-world Phase measurement unit data using the local outlier factor algorithm and Density-Based Spatial Clustering of Application with Noise (DBSCAN). The mean absolute error, root mean square error and R^2-score criteria are used to validate the proposed method's data filling results against contemporary techniques such as decision tree and linear regression algorithms.

Keywords: data cleaning, outlier detection (OD), data recovery, phase measurement unit(PMU), apache spark

OPEN ACCESS

Edited by:
Dazhong Ma,
Northeastern University, China

Reviewed by:
Liang Chen,
Nanjing University of Information Science and Technology, China
Biplob Ray,
Central Queensland University, Australia

****Correspondence:***
Risheng Qin
qinrisheng2020@126.com

Specialty section:
This article was submitted to Smart Grids, a section of the journal Frontiers in Energy Research

Received: *14 April 2021*
Accepted: *16 July 2021*
Published: *27 July 2021*

Citation:
Shen L, He X, Liu M, Qin R, Guo C, Meng X and Duan R (2021) A Flexible Ensemble Algorithm for Big Data Cleaning of PMUs. Front. Energy Res. 9:695057. doi: 10.3389/fenrg.2021.695057

INTRODUCTION

Due to the increasing demand for accurate control and management in smart grids, many advanced online monitoring devices have been installed and provide abundant operating data resources using Phase Measurement Units (PMUs). The data preprocessing is an important step that transforms the raw operating data used in the load forecasting model, user clustering tool, equipment maintenance, and energy theft detection technique. The outcome of data preprocessing has a significant impact on the data modelling process. For instance, a prediction model fed by a raw dataset with noise and bad data will be inefficient and cause inaccuracy. PMU failures, such as communication errors and noises, cause irregular packet data and asymmetric magnitude spikes, which are particularly problematic for smart grid applications. As a result, PMUs' data cleaning algorithm must maintain high speed and sensitivity to faulty data in order to deliver a highly reliable data mining model. However, designing a data cleaning algorithm that balances high speed and sensitivity is a technological challenge that needs to be addressed.

Data cleaning technologies are a heavily studied domain of data statistics and machine learning. The whole process of extensive data cleaning is illustrated as outlier detection and data filling. The outliers which do not follow the main of the data may be produced by inducing random errors and faulty measurements (Zhao et al., 2019). For outlier detection, with the recent advancement in machine learning techniques, both unsupervised and supervised methods have been investigated for

better accuracy, speed, and computation cost. In supervised models, such as one-class support vector machine (SVM) (Ma and Perkins, 2003), decision forest (Reif et al., 2008), convolutional neural network (Ren et al., 2020), and the long short-term memory network (LSTM) (Wu et al., 2020) can achieve excellent performance by learning massive labeled data. However, labeling massive data is very time-consuming and needs great manual effort, which limits its application at an industrial scale. In comparison, unsupervised outlier detection does not need labeling and can achieve good accuracy in most cases. Even though some of their results are poor in complicated scenarios, unsupervised methods, namely Kmeans, Gaussian Mixture Model (GMM), CURE (Lathiya and Rani, 2016), Density-Based Spatial Clustering of Application with Noise (DBSCAN) (Manh and Kim, 2011), local outlier factor (LOF) (Pokrajac et al., 2007) and isolation forest (iForest) (Liu et al., 2008) are extensively used in real-world scenarios because they are easy to implement. Subsequently, there have been several attempts to use an unsupervised model to clean PMU data in the smart grid. For example, in (Mahapatra et al., 2016), principal component analysis (PCA) is used to detect outliers in PMU measurements. Likewise, PCA is incorporated with an artificial neural network (ANN) to improve detection accuracy (Mahapatra et al., 2017).

Meanwhile, researchers have been drawn to the drawbacks of stand-alone approaches, which produce inconsistent results in complex situations. As a result, various ensemble-based models have been designed to address deficiencies in real-world applications and improve their performance. For example, to improve accuracy, the local outlier factor (LOF) algorithm, correlation outlier probabilities, and single-linkage-based outlier detection methods are used (Kummerow et al., 2018). The DBSCAN, Chebyshev, and linear regression models are combined to predict PMU outliers (Zhou et al., 2019), but the approach cannot distinguish abnormal and regular operations. The Kmeans and local outlier probability methods are used to identify various types of anomalies based on the iForest anomaly score, such as fault detection, transient disturbance, etc. (Khaledian et al., 2020). In complex scenarios, these ensemble methods can present improved performance. However, the performance of extremely big data sets that may be computationally analyzed to discover patterns is rarely mentioned.

With the development and deployment of PMUs, the size of received data risen exponentially for a data center. (Khan et al., 2014; Yang et al., 2015). When dealing with vast amounts of data, conventional data processing methods can take days or weeks, which is insufficient time for data analysis. As a result, to ensure successful data processing, some attempts focus on big data technology. An adaptive hoeffding tree with a transfer learning approach is proposed (Mrabet et al., 2019) to detect the PMU data's events. In another attempt, a feature generation system is well-designed via Apache Spark core, which successfully fits 400 PMUs from the North American power grid (Kumar et al., 2021). A streaming interface based on Apache Spark for the synchrophasor data stream is investigated (Menon et al., 2018). Despite this, the integration and expansion of detection algorithms on existing big data platforms have limitations.

Furthermore, data filling is often addressed in publications as an important step in avoiding missing values. Statistic techniques and machine learning methods can complete the data filling processing. For statistic techniques, an improved cubic spline interpolation method is used to recover the missing data in the transient state and static state of power systems (Yang et al., 2019). A feature component extraction-based approach is proposed to recover a single channel data of PMU, which accounts for more details of the data waveform (Gao et al., 2016), but the relationship between PMUs is ignored. By contrast, an extreme learning machine and a random vector functional link model are introduced to produce good filling results (Li et al., 2019). Besides, artificial neural network technologies are also developed to achieve a good performance against complex scenarios. For example, a least-squares generative adversarial network is adopted to generate adequate monitoring data (Wang et al., 2021). Except for developing a new method, the researchers utilize the potential information in power systems' features to improve the accuracy, such as network topologies and operation mode. In (Ren and Xu, 2019), the network topologies are considered in a recovery program based on a generative adversarial network (GAN). Although the importance of topology in data recovery processing is investigated, publications seldom cover the whole data cleaning process, including outlier detection and data recovery.

Traditional bad data detection algorithms may underperform when dealing with complex scenarios and take a long time to run without big data technologies. Our motivation is to investigate how to apply the complete data cleaning process of PMUs, including outlier detection and data filling, to existing big data platforms to achieve expected performance. A flexible ensemble approach for data cleaning is given in this study to adapt to the failure of a single technique. In outliers detection, we adopt an ensemble method that includes three sub-detectors, the Kmeans combined with PCA, GMM, and iForest. A flexible voting mechanism then aggregates their results, and the aggregation is used to label outliers. After the outliers detection, the Gradient Boost Decision Tree (GBDT) is used and well designed to recover missing data and observed outliers. Apache Spark platform, Spark streaming system, Kafka and Hadoop distributed file system is selected to perform and test the proposed algorithm with massive datasets. In more detail, the contributions of this paper are listed as follows. First, a flexible data cleaning algorithm uses Apache Spark to automate the identification of outliers and retrieve missing data. Second, we propose a flexible voting mechanism for outlier detection to aggregate the outputs of PCA-Kmeans, GMM, and iForest in complex cleaning scenarios.

PROBLEM DESCRIPTIONS

The Framework of Proposed Data Cleaning via Spark

Figure 1 depicts a hierarchical data-cleaning framework proposed in this paper. The presented data cleaning algorithm is deployed in the Spark and Hadoop distributed file systems.

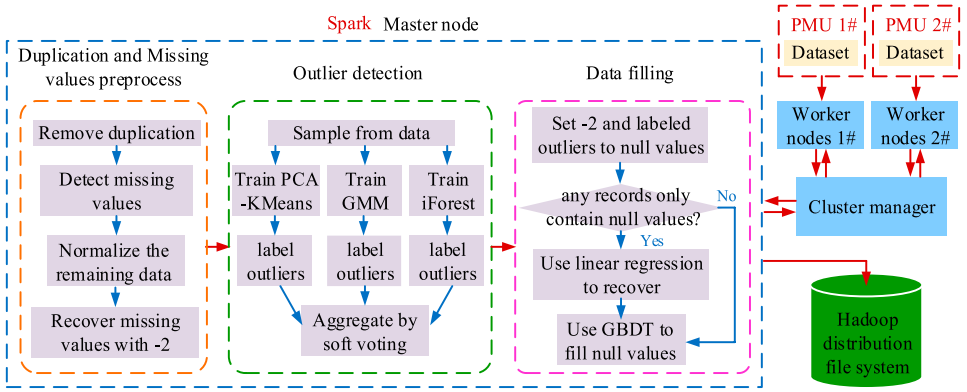

FIGURE 1 | The framework of data cleaning in Spark.

Master nodes and worker nodes are included in the system (2 nodes, as shown in **Figure 1**). When the proposed algorithm interacts with the master node, the master node asks the cluster manager for computing resources. The cluster manager responds by allocating jobs to worker nodes, and the worker nodes perform tasks based on PMU data.

The proposed data cleaning process is divided into three stages, as shown in **Figure 1**. The first step is to prepare the data. In this stage, the PMU data is uploaded to the worker nodes, preparing them for the next cleaning process. The cleaning procedure is preceded by a preprocess duplication and missing values. We remove duplication data and then find missing values. In this condition, the remaining data with noises are normalized. After that, the missing values labelled with "−2" and the normalized data are combined to form the dataset. Choosing −2 is to distinguish missing values from the normalized data (Liu et al., 2020). In the second stage, we randomly sample from the dataset to train PCA-KMeans, GMM, and iForest algorithms to predict outliers using a soft voting mechanism. Note that outliers include noise data. In the third stage, the outliers and missing values with "−2" are replaced with null values due to their abnormal features. If any record only contains null values, linear regression is used to recover this record.

Outlier Detection

In general, outlier detection algorithms should be unresponsive to normal data, resilient and robust to outliers, and capable of computation. However, only a few algorithms can meet the requirement in most cases, and the algorithm's output can jeopardize the data analysis credibility. To be more specific, 1) the algorithm may be insensitive to one or more types of outliers, such as bad data or missing values. 2) the model with adjustable parameters generates a high computational cost when cleaning a large dataset and can result in overfitting. 3) The algorithm may be vulnerable to standard power system manual operations, such as network topology changes.

To demonstrate more clearly, we take a section of PMU data shown in **Figure 2**, where five points are identified as outliers and highlighted in the figure. The state-of-the-art detecting algorithms, Kmeans, GMM, iForest, DBSCAN, and LOF, are

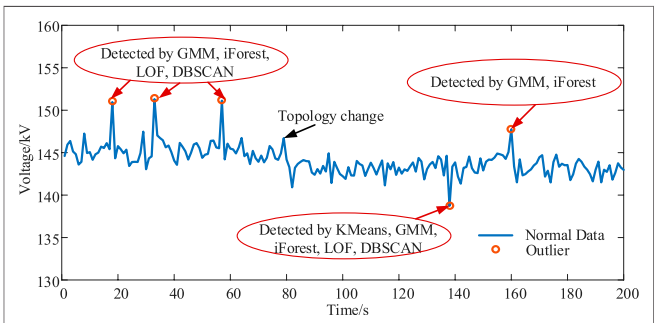

FIGURE 2 | The framework of data cleaning in Spark.

compared, with their parameters tuned. Most of the algorithms miss two outliers due to the topology change. However, only a small number of algorithms are capable of detecting all outliers.

To overcome these challenges, combining different findings from different detectors is necessary. The combining model can take advantage of every detector by aggregation and coherently achieve better performance. Its aggregation mechanism is the key to utilize the benefits fully. This paper investigates a flexible voting aggregation mechanism for the ensemble method to identify outliers.

Furthermore, in an ensemble algorithm, sub-detector selection is a critical step. In theory, any outlier detectors can be used for the ensemble, but since the compute resource is limited, the sub-detector number is limited. In the sub-detector selection, the detectors based on different methodologies are welcomed. In this paper, the density-based method, iForest, is chosen because of its high scalability and low memory use. The clustering-based methods, Kmeans and GMM, are used since the Kmeans ease of implementation in distributed computing. The GMM is selected because of its fuzzy clustering, which provides the probability of data points belonging to each cluster and is more flexible than Kmeans. While starting the cleaning process, three detectors are trained by the sampling data and then process the entire data separately and simultaneously using Spark's pipeline mechanism which can improve computing efficiency.

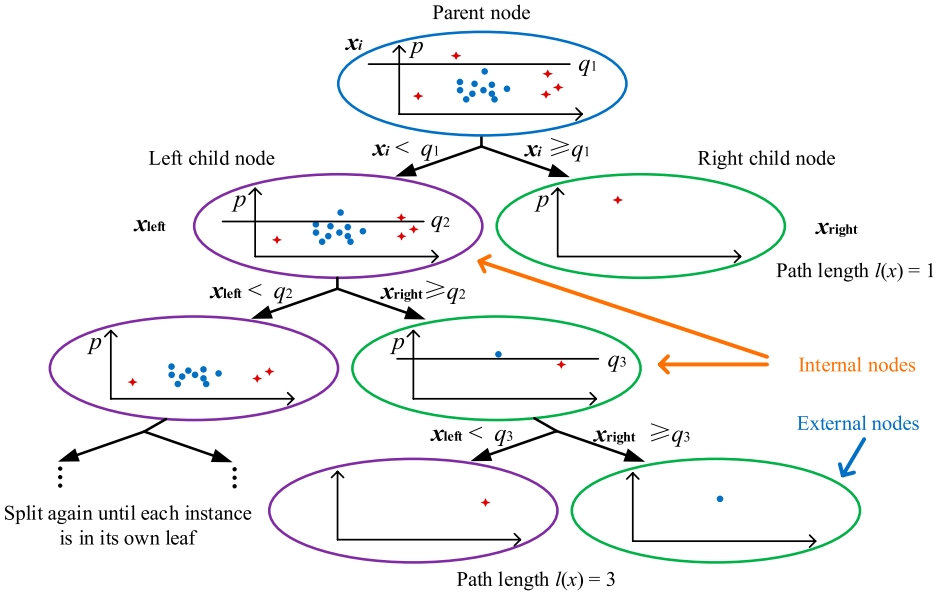

FIGURE 3 | The mechanism of iTree.

Data Filling

Standard manual operations, such as network topology changes and line maintenance, often occur and cause PMU data to drift. Some filling algorithms, on the other hand, ignore the information and predict a significant error. As a result, such information should be considered when training a filling algorithm.

Furthermore, the filling algorithm's accuracy should be given more consideration. As a famous filling algorithm, GBDT can reach a high accuracy than other filling algorithms. The GBDT is a classic ensemble learning method that creates a strong regression tree by combining weak regression trees (typically train classification and regression tree (CART)). Thereupon, GBDT handles nonlinear relationships well and achieves high accuracy in fragmented datasets. Therefore, we adopt the GBDT method against missing data packets.

ENSEMBLE MODELING FOR OUTLIER DETECTION

Data Preparation

In this subsection, an ensemble method based on sub-detector PCA-Kmeans, GMM, and the iForest algorithm is proposed in order to obtain a more accurate detection of an outlier. To clearly illustrate the process, let $D = d_k, d_{k+1}, d_{k+2}, \ldots, d_{k+w}$ be the kth data window with size w, while D is a set of data rows. In which each data row $d_i$contains seven components: voltage magnitude, current magnitude, current angle, active power, apparent power, reactive power, and power factor angle.

PCA-Kmeans Detector

The Kmeans is a classical classifying method that marks the data into several clusters. By analyzing and classifying the clusters, the clusters of outliers can be detected. However, given the potential vulnerability of the Kmeans on high dimensional data, the PCA approach is combined with Kmeans to reduce the dimension of the data, called the PCA-Kmeans detector. The PCA approach is one of the most popular dimensionality reduction techniques (Mahapatra et al., 2017), aiming to find an orthogonal subspace whose basis vectors correspond to the maximum-variance directions in the original space. By using the output of the PCA model, the Kmeans method can achieve better accuracy. For clarity, let take $B = \{b_1, b_2, \ldots, b_i, \ldots, b_w\}$ as the output of PCA. Each b_i has n_{sub} features. In the Kmeans method, each $b_i$of B should be assigned to the cluster which has the least squared Euclidean distance (Khaledian et al., 2020). To begin with, the k number of the centroid is selected randomly as $m_1^{(1)}, \ldots, m_k^{(1)}$. Whereas a centroid is a data point at the cluster center. Next, iterations are implemented to find the nearest centroid for each b_i, as given by **Eq. 1**.

$$C_i^{(t)} = \left\{ b_i : \| b_i - m_i^{(t)} \|^2 \leq \| b_i - m_j^{(t)} \|^2 \, \forall j, 1 \leq j \leq n \right\} \tag{1}$$

where $C_i^{(t)}$ is the serial number of the nearest cluster for b_i in t iteration, $1 \leq C_i^{(t)} \leq n$. $m_i^{(t)}$ is the mass point of $C_i^{(t)}$. n is the number of clusters.

After labeling each feature set in every iteration, the centroid in each step will be updated by **Eq. 2**.

$$m_i^{(t+1)} = \frac{1}{\left| C_i^{(t)} \right|} \sum_{b_i \in C_i^{(t)}} b_i \tag{2}$$

Meanwhile, when the centroid difference in an adjacent iteration is less than ξ, the iteration comes to a halt and gives final labels to each feature set in vector B based on **Eq. 3**.

$$m_i^{(t+1)} - m_i^{(t+1)} \leq \xi \tag{3}$$

where E_i is the mean of data points in C_i , ξ is a very small positive number. Here, we take the result of the PCA-Kmeans that is a set of cluster labels S_{kmeans}.

Gaussian Mixture Model-Based Detector

The GMM is a useful algorithm for detecting outliers based on a density function (De la Torre et al., 2012). Unlike Kmeans, the data is assumed to be modelled by several Gaussian density functions in this method. Each Gaussian density in the kth is given by a Gaussian function **Eq. 4**. The GMM model is the weighted sum of several Gaussian densities, illustrated by **Eq. 5**.

$$g_k(d|\mu_k, \sigma_k) = \frac{1}{\sqrt{2\pi\sigma_k^2}} e^{-\frac{(d-\mu_k)^2}{2\sigma_k^2}} \tag{4}$$

$$g(D; \mu, \pi, \sigma) = \sum_{k=1}^{M} \pi_k g_k(D; \mu_k, \sigma_k) \tag{5}$$

where $\pi = \{\pi_1, \ldots, \pi_M\}$, $\mu = \{\mu_1, \ldots, \mu_M\}$, $\sigma = \{\sigma_1, \ldots, \sigma_M\}$, and $\pi_k$is the specific weights of each Gaussian model within a mixture. M is the number of Gaussian function. μ_k, $\sigma_k$are the means and the covariance matrix of each model, respectively.

To determine the parameters such as π, μ, σ of the Gaussian functions, the maximum likelihood function given by **Eq. 6** is used for help by using the Expectation-Maximization (EM Algorithm) (De la Torre et al., 2012). The log-likelihood is used as **Eq. 7** to determine if a data point belongs to the Gaussian functions measured earlier. The GMM's output is then assigned the weight of each data point to simple Gaussian density.

$$L(\mu, \pi, \sigma) = \prod_{i=1}^{w} \sum_{k=1}^{M} \pi_k g_k(d_i; \mu_k, \sigma_k) \tag{6}$$

$$\log L(\mu, \pi, \sigma) = \sum_{k=1}^{M} \sum_{i=1}^{w} z_{ki} \{\log(\pi_k) + \log(g_k(d_i; \mu_k, \sigma_k))\} \tag{7}$$

Z_{ki} contains 0 or 1 depending on whether the data d_i belongs to Gaussian function k.

The mean log-likelihood criterion is then used to determine if the incoming data in the next window matches with the current GMM or not (Diaz-Rozo et al., 2018); it is calculated using **Eq. 8**.

$$\overline{\log L}(\mu, \pi, \sigma) = \frac{1}{w} \sum_{i=1}^{w} \log\left(\sum_{k=1}^{M} \pi_k g_k(d; \mu_k, \sigma_k) \right) \tag{8}$$

Isolation Forest Detector

In general, anomalies are less common than normal findings and have different values. The Isolation Forest algorithm takes advantage of this feature to measure a dataset's anomaly ratings, which are then used to distinguish outlier points (Liu et al., 2008). In this subsection, isolation trees (iTree) and path lengths are introduced.

For clarification, let us take a random binary tree as an example; partitioning observations is repeated recursively until all the observations are isolated. As shown in **Figure 3**, the iTree that uses a binary tree structure is proposed to isolate observations.

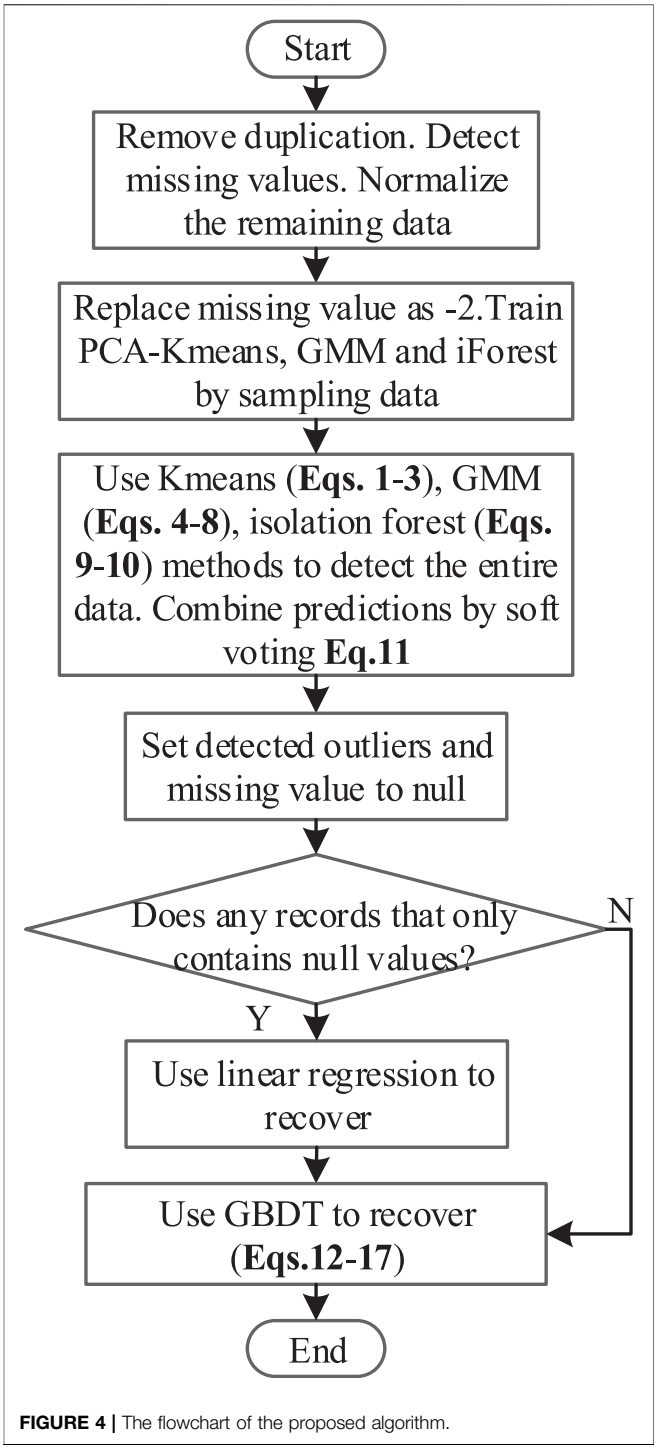

FIGURE 4 | The flowchart of the proposed algorithm.

Definition 1 (iTree): iTree is a random binary tree with no more than two children per node. As shown in **Figure 3**, internal nodes have exactly two children, while external nodes have none. Each internal node has a randomly chosen function q and a split value p, resulting in the node's split into two child nodes according to the condition $q < p$. This process is repeated until all of the nodes have just one case.

We denote a training dataset with N instances by $X = \{x_1, \ldots, x_N\}$. The subsampled set $\mathcal{X} \subset X$ is sampled from X with φ instances, which is utilized for training an iTree. The process of building an iTree is to divide up the subsampled set $\mathcal{X}$ recursively into subspaces. Note that we adopt only subsampled sets of small fixed sizes to build iTrees, regardless of the dataset's size. This way, we can obtain each iTree very swiftly.

Anomalies are isolated closer to the root node of an iTree and have short path lengths, as seen in **Figure 3**. On the other hand, standard points are isolated at the deep end of an iTree and therefore have long path lengths. As a result, anomaly scores are a function of path lengths. The length of the route is determined as follows.

Definition 2 (Path length): $l(x)$ is the number of edges between the root node and the external node corresponding to an instance x in the iTree.

For the same dataset X, we can build multiple iTrees that are constructed by randomly selected features, split values, and subsampled datasets. To aggregate the results of iTrees and calculate the anomaly score, we first introduce an average path length $c(\varphi)$ for instances φ in an iTree calculated by **Eq. 9**. This average path length can represent the length situation of the instances φ, which is used to normalize the length of each component x in the instances. Next, the anomaly score of each component x in the instance φ can be obtained by calculating **Eq. 10**. The anomaly score ranges from 0 to 1, and the data instance will be normal if the score is lower than 0.5 (Liu et al., 2008). Further, the data instance which is closed to 1 can be detected as an outlier.

$$\begin{cases} c(\boldsymbol{\varphi}) = 2H(\boldsymbol{\varphi}-1) - \dfrac{2(\boldsymbol{\varphi}-1)}{\boldsymbol{\varphi}} \\ H(\boldsymbol{\varphi}-1) \approx log(\boldsymbol{\varphi}-1) + e \end{cases} \tag{9}$$

$$\begin{cases} s(\boldsymbol{x}, \boldsymbol{\varphi}) = 0.5 - 2^{-\frac{E(l(x))}{c(\varphi)}} \\ E(\boldsymbol{l}(\boldsymbol{x})) = \dfrac{1}{\boldsymbol{N}_{Tree}} \sum_{n=1}^{N_{Tree}} \boldsymbol{l}_n(\boldsymbol{x}) \end{cases} \tag{10}$$

where e is the Euler constant; $l(x)$ is the path length of each component x in the instance φ. The expected path length is represented as $E(l(x))$. N_{Tree} is the number of iTrees.

Soft Voting Mechanism

To fully utilize the advantages of sub-detectors, a soft voting mechanism is used to combine the sub-detectors predictions and increase robustness to complex scenarios. In particular, compared to the outlier probability given by GMM and iForest, the prediction of Kmeans is "hard" and has less elasticity against the scenario because it only gives a cluster label to each data point. The Kmeans prediction should be combined with another "soft" approach with a similar mechanism to deal with the poor results. For example, GMM, a soft clustering method, is used to multiple the Kmeans results marked as $\boldsymbol{S}_{kmeans}\boldsymbol{P}_{GMM}$. Although $S_{\text{Kmeans}} = 1$, which means outlier detected in the Kmeans method, the outlier probability is still driven by GMM. Furthermore, to account for diversity in our voting mechanism's final prediction, the average outlier likelihood of all sub-detectors is used, as seen in **Eq. 11**.

TABLE 1 | The details of outlier detection datasets.

Dataset	Points	Feature	Outliers (%)
Satellite	6,435	36	2036 (32%)
Shuttle	49,097	9	3,511 (7%)
BreastW	683	9	239 (35%)
Http	567,479	3	2,211 (0.4%)

TABLE 2 | The results of the proposed method with outlier detection datasets.

Metrics	Datasets			
	Satellite	Shuttle	BreastW	Http
T_P	1,343	3,375	222	2,185
F_P	145	145	17	84
T_N	4,654	45,440	427	565,203
F_N	293	137	17	26
Precision	0.90	0.96	0.93	0.96
Recall	0.82	0.96	0.93	0.99

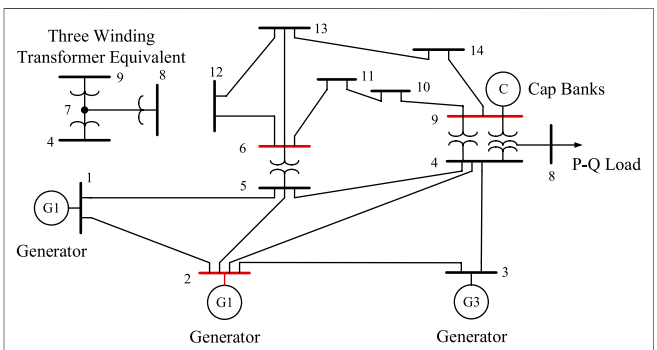

FIGURE 5 | The testing IEEE 14-bus system.

$$\boldsymbol{P} = \frac{\boldsymbol{S}_{kmeans}\boldsymbol{P}_{GMM} + \boldsymbol{P}_{GMM} + \boldsymbol{P}_{iForest}}{3} \tag{11}$$

SKmeans, PGMM, and PiForest are the output of the PCA-KMeans, GMM, and iForest algorithms. S_{Kmeans} is a binary variable, and $S_{\text{Kmeans}} = 0$ addresses the normal data, while abnormal data is annotated as 1. P_{GMM} is the probability of outliers for an observation, which is closed to 1, meaning outlier. P_{iForest} is the anomaly score of the data point.

DATA FILLING PROCESS AND DATA CLEANING FUNCTION

Gradient Boosting Decision Tree-Based Filler

As discussed in *Problem Descriptions*, data filling is an important part of data cleaning, and it is a regression problem by definition. For PMU data, it is possible to have missing values for each feature, which presents as single or continuous types in a dataset. To tackle different types of missing values, the GBDT model is trained for each feature of PMU data, respectively. In case of single missing value occurs in a feature, the GBDT model can

TABLE 3 | The noises injection of simulated scenarios.

Abnormal type	Anomalies calculation in each feature	Abnormal/Data points
Noise 5%	1.p.u *105% + $G(x)$	569/4,000
Noise 10%	1.p.u *105% + $G(x)$	1,091/4,000
Noise 15%	1.p.u *105% + $G(x)$	1,529/4,000

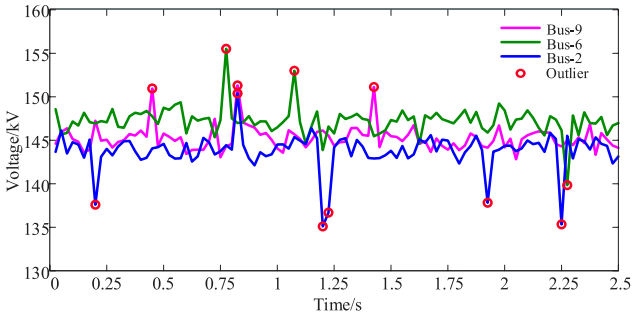

FIGURE 6 | The segment of synthetic noise-5% PMU data.

TABLE 4 | The outlier detection results in simulated scenarios.

Algorithm	Metrics	Abnormal rate		
		5%	10%	15%
Proposed FEA	T_P	564	1,077	1,510
	F_P	16	14	21
	T_N	3,415	2,895	2,397
	F_N	5	14	19
	Precision	0.972	0.987	0.986
	Recall	0.991	0.987	0.987
DBSCAN	T_P	568	1,090	1,529
	F_P	6	15	0
	T_N	3,425	2,856	2,430
	F_N	1	38	41
	Precision	0.989	0.986	1
	Recall	0.998	0.966	0.974
LOF	T_P	550	1,047	1,484
	F_P	19	44	45
	T_N	3,412	2,891	2,455
	F_N	19	18	16
	Precision	0.97	0.960	0.970
	Recall	0.97	0.983	0.989

easily fill it using the other features as input. By contrast, when facing the continuous missing values loss of all features, the topology is the first to be recovered using the last instance. Then, the variables strongly associated with time—such as active power—are recovered by the linear regression method. Next, the other features are retrieved by the GBDT method.

The GBDT is used as a filler and to model an approximation function $f(X)$ of a specified result $Y = \{y_1, y_2, \ldots, y_n\}$ with a set of the input variable as $X = \{x_1, x_2, \ldots, x_{n_{sp}}\}$. n_{sp} is the length. During the approximation process, a loss function is usually adopted to search for the most precise approximation function. As illustrated in **Eq. 12**, the most precise model is obtained when the loss function is minimum. Here, we select the squared error function as the loss function shown in **Eq. 13**.

$$\boldsymbol{F}(\boldsymbol{x}) = \arg\min_{f(x)} \boldsymbol{L}(\boldsymbol{y}, \boldsymbol{f}(\boldsymbol{x})) \tag{12}$$

$$\boldsymbol{L}(\boldsymbol{y}, \boldsymbol{f}(\boldsymbol{x})) = [\boldsymbol{y} - \boldsymbol{f}(\boldsymbol{x})]^2 \tag{13}$$

The optimization can be effectively solved by a gradient descent algorithm, and the approximation function can be updated using the results of every iteration, illustrated via **Eq. 14**. In each iteration, the GBDT model uses the results in the last iteration and a classification and regression tree (CART), which is updated as **Eq. 14**. Especially in the initial iteration, $f_0(x) = 0$.

$$\begin{cases} f(\boldsymbol{x}) = \sum_{m=1}^{M} \boldsymbol{f}_m(\boldsymbol{x}) \\ \boldsymbol{f}_m(\boldsymbol{x}) = \boldsymbol{f}_{m-1}(\boldsymbol{x}) + \boldsymbol{\gamma}_m \sum_{j=1}^{J} \boldsymbol{c}_{mj} \boldsymbol{I}, \quad x \in \boldsymbol{R}_{mJ} \end{cases} \tag{14}$$

where M is the length of iterations. m is the serial number of iteration. $\sum_{j=1}^{J} \boldsymbol{c}_{mj}\boldsymbol{I}$ is the result of the CART. J is the number of leaf nodes of the CART. The area disjointed by each leaf node is R_{m1}, R_{m2}, ..., R_{mJ}. c_{mj} is the prediction value of jth area. $\boldsymbol{\gamma}_m$can be calculated by **Eq. 15**, and y_i is the actual value of variable y.

$$\boldsymbol{\gamma}_m = \arg\min_{\gamma} \sum_{i=1} \boldsymbol{L}\left(\boldsymbol{y}_i, \boldsymbol{f}_{m-1}(\boldsymbol{x}_i) + \boldsymbol{\gamma} \sum_{j=1}^{J} \boldsymbol{c}_{mj} \boldsymbol{I},\right) \tag{15}$$

By repeating the above interactive steps, the output of GBDT can be obtained by the final iteration.

The Proposed Processing of Data Cleaning

A flowchart of the proposed strategy is shown in **Figure 4**. Step 1: after eliminating duplication and detecting missing values, normalize the remaining data. Step 2: replace missing values with '-2' and train PCA-KMeans, GMM, and iForest algorithms by sampling the normalized data. Step 3: detect the entire data by **Eqs. 1–10** and combine PCA-KMeans, GMM, and iForest to eliminate outliers via a soft voting approach. Step 4: if any record only contains null values, using linear regression recovers the time-dependent features of records and then employing GBDT recovers the entire data. Step 5: otherwise, GBDT is used to recover the entire data.

NUMERICAL SIMULATION

Experimental Settings

In this simulation, the detailed experimental evaluation is presented with Spark 2.4.0, Kafka 0.10.1.0, Hadoop 2.4.7 under Ubuntu 16.04 operation system. Three scenarios are presented to demonstrate the feasibility of the proposed process. The outlier

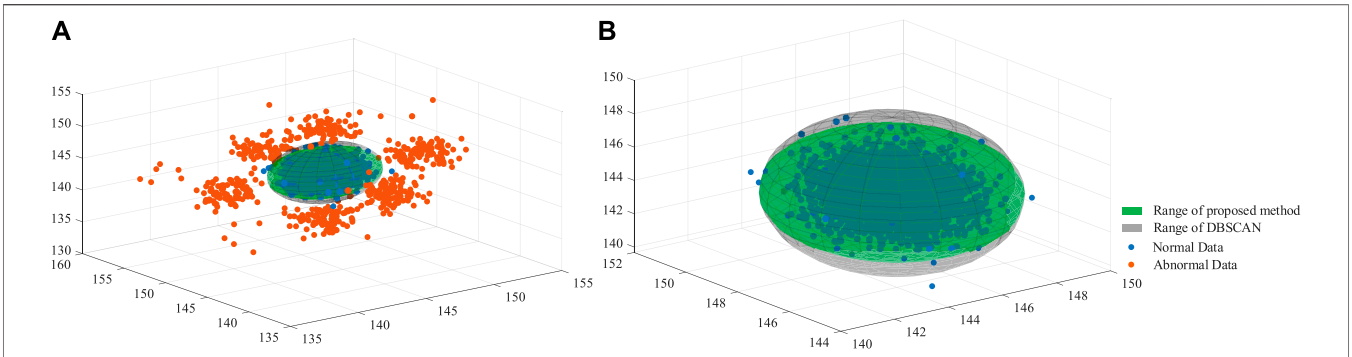

FIGURE 7 | (A) The distribution of normal and abnormal data in noise-5% **(B)** The range of proposed FEA and DBSCAN.

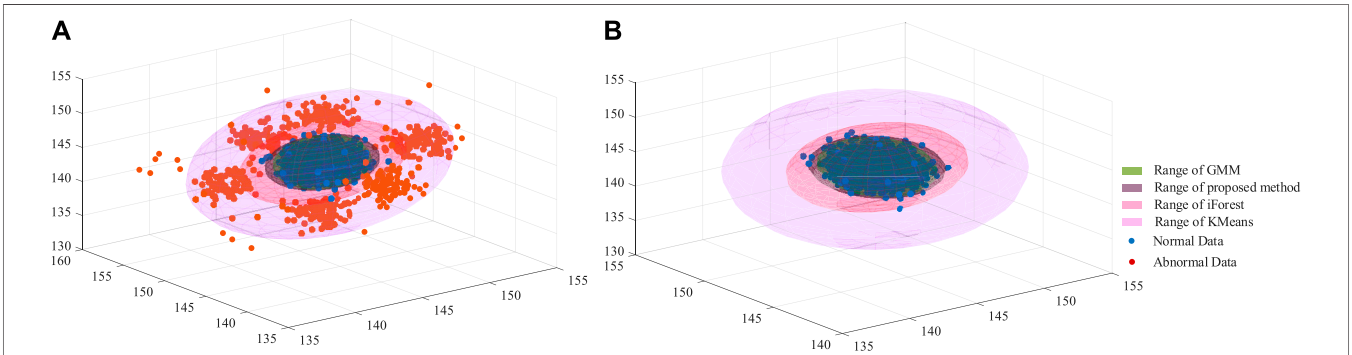

FIGURE 8 | (A) The distribution of noise-5% data with a range of different detectors **(B)** The range of proposed FEA and GMM, KMeans and iForest.

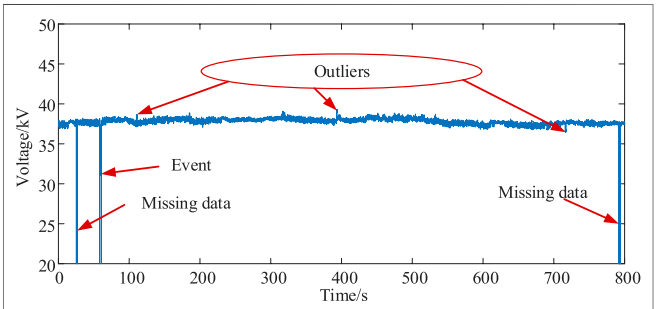

FIGURE 9 | A feature of real-world PMU data.

TABLE 5 | The outlier detection results in real-world scenarios.

Algorithm	Metrics					
	T_P	F_P	T_N	F_N	Precision	Recall
FEA	1,077	10	33,811	46	0.991	0.959
DBSCAN	1,082	71	33,769	22	0.982	0.964
LOF	1,102	51	33,740	51	0.956	0.956

identification function of the proposed approach is firstly evaluated by an industrial dataset from the reference (Liu et al., 2008), considering precision and recall metrics. Secondly, the outlier detection function is examined using

TABLE 6 | The data recovery results in real-world scenarios.

Algorithm	Metrics		
	MAE	*RMSE*	R^2-*score*
FEA-GBDT	0.354	0.463	0.910
Decision tree	0.366	0.483	0.904
Linear regression	0.378	0.476	0.900

simulated PMU data and real PMU data. Finally, the mean absolute error and the root mean squared error are employed to evaluate the precision of the proposed approach in recovering data compared with the linear regression algorithm and the decision tree approach.

Outlier Detection of the Public Industrial Dataset

The proposed algorithm (FEA) is recommended in this scenario for detecting real-world datasets from outlier detection datasets and generating a score. Considering Satellite, Shuttle, Breastw, and Http datasets (Liu et al., 2008) illustrated in **Table 1**, a confusion matrix, which includes false positive (F_p), false negative (F_n), true positive (T_p), and true negative (T_n), is used to validate the performance of the proposed algorithm. Following that, we can use **Eqs. 18**, **19** to measure the recall and precision ratios for further discussion.

$$\boldsymbol{recall} = \frac{\boldsymbol{T_p}}{\boldsymbol{T_p} + \boldsymbol{T_N}} \tag{18}$$

$$\boldsymbol{precision} = \frac{\boldsymbol{T_p}}{\boldsymbol{T_p} + \boldsymbol{F_P}} \tag{19}$$

The number of outliers detected as outliers is T_P, and the number of normal data detected as normal data is T_N. At the same time, F_P stands for the number of normal data points that have been identified as outliers. The number of outliers detected as normal data is given by F_N.

As shown in **Table 2**, the proposed FEA can achieve good performance while cleaning all types of data with large and highly polluted information, although this recall is about 82% for Satellite.

Outlier Detection of Synthetic PMU Dataset

Using PMU operational data, the proposed method and other methods are compared in this subsection. In PSCAD/EMTDC, simulation data is produced using a model IEEE 14-bus network system with PMUs installed on bus-2,6,9, as shown in **Figure 5**. The length of operation data of PMU is 4,000 points with a sampling rate of 40 frames per second. The data is polluted by outliers and missing values using a Gaussian-distributed random function as $z = G(x)$. **Table 3** shows that each PMU data has 5 percent -15 percent noise and 5 percent -15 percent missing values injected into it. As an example, if a data point has a voltage feature of 35kV, the noise is calculated as 35*105% + $G(x)$. As illustrated in **Figure 6**, one segment of the synthetic data is added by a noise-5% PMU data.

Changing the ratios of white noises and null values, the proposed FEA can maintain an expected performance, as shown in **Table 4**. For instance, considering the dataset with 15% noise, DBSCAN and FEA have similar results. As illustrated in **Figure 7A,** the normal and abnormal data are used to predict the outliers. As shown in **Figure 7B**, DBSCAN has a little larger cover of normal data than FEA does, which means that DBSCAN can achieve slightly better precision than FEA.

Following that, **Figure 8A** illustrates the range of sub-detectors used in our ensemble method. The details indicate that the KMeans range is the largest but includes some abnormal data, indicating that this method detects more outliers than normal data (FN), as shown in **Figure 8B** iForest has a smaller range than KMeans but perfectly covers all normal data. GMM has a narrow range and may predict more normal data as outliers (FP). By combining the advantages of each sub-detector, the FEA can achieve a normal range size while maintaining a high level of outlier detection performance.

Outlier Detection of Real-World PMU Dataset

For performance estimation, real-world PMU data from a specific region in southwest China is used, and domain experts label outliers and missing values in the dataset. FEA can detect outliers and missing values, as shown in **Figure 9**. As presented in **Table 5**, the FEA can efficiently clean data with a precision of 99.1 percent and a recall of 95.9 percent. The good performances in real-world PMU data again verify the proposed FEA's effectiveness.

Data Recovery of Real-World PMU Dataset

The linear regression, decision tree, and GBDT algorithms are introduced in this sub-section to complete the regression training process and fill null values with real-world PMU data. The root squared measurement error (RSME), mean absolute error (MAE), and R^2-score are respectively calculated to evaluate the performance of the proposed approach in **Eqs. 20–22**.

$$\boldsymbol{MAE} = \frac{1}{\boldsymbol{N}} \sum_{\boldsymbol{i}=1}^{\boldsymbol{N}} \left| \boldsymbol{y_i} - \boldsymbol{h}(\boldsymbol{x_i}) \right| \tag{20}$$

$$\boldsymbol{RSME} = \sqrt{\frac{1}{\boldsymbol{N}} \sum_{\boldsymbol{i}=1}^{\boldsymbol{N}} \left(\boldsymbol{y_i} - \boldsymbol{h}(\boldsymbol{x_i}) \right)^2} \tag{21}$$

$$\boldsymbol{R^2} = 1 - \frac{\sum_{i=1}^{N} \left(\boldsymbol{y_i} - \boldsymbol{h}(\boldsymbol{x_i}) \right)^2}{\sum_{i=1}^{N} \left(\boldsymbol{y_i} - \overline{\boldsymbol{y_i}} \right)^2} \tag{22}$$

where N is the size of data, y_i is data point, $H(x_i)$ is the prediction with the input x_i, and $\overline{y_i}$ is the average of data.

As illustrated in **Table 6**, the performance of the proposed FEA-GBDT is superior to that of the other algorithms because of lower MAE and RMSE and larger R^2-score.

CONCLUSION

This paper proposes a modular ensemble-based cleaning approach for PMUs to achieve outlier detection and data filling using big data technologies. The proposed approach considers and aggregates the advantages of different methods such as KMeans, GMM, and iForest for outlier identification, allowing it to perform better. Missing values due to system error are also investigated and retrieved using the proposed process. Notably, computational results show that the proposed approach can effectively process outliers, is resilient to a high percentage of bad data, and performs well with a large dataset. The proposed method achieves accurate prediction as compared to DBSCAN and LOF algorithms. The proposed approach, in particular, can handle large datasets deployed on Hadoop and Spark systems. When data filling is taken into account, our model produces a lower mean absolute error and root squared measurement error and R^2-score. Furthermore, our algorithm results show that using big data technology, a single detector's poor performance and low efficiency can be replaced by a high-efficiency ensemble approach. PMUs' outlier detection and data filling functions have the potential to clean and use data in real-time for fault detection, data processing, and prediction.

Some factors, such as communication infrastructure and system maintenance, may have an impact on the proposed algorithm's efficiency. As a result, our future work will focus on taking into account the aforementioned considerations and refining the proposed approach in these scenarios.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

AUTHOR CONTRIBUTIONS

Conception and design of study: LS Acquisition of data: ML, XH Drafting the article: RQ, XH Analysis and interpretation of data: LS, RQ, ML, CG Revising the article critically for important intellectual content: RQ, XM, RD.

FUNDING

This work was supported by the Science and Technology Foundation of China Southern Power Grid (YNKJXM20191369).

REFERENCES

De la Torre, F. (2012). A Least-Squares Framework for Component Analysis. *IEEE Trans. Pattern Anal. Mach. Intell.* 34 (6), 1041–1055. doi:10.1109/TPAMI.2011.184

Diaz-Rozo, J., Bielza, C., and Larrañaga, P. (2018). Clustering of Data Streams with Dynamic Gaussian Mixture Models: An IoT Application in Industrial Processes. *IEEE Internet Things J.* 5 (5), 3533–3547. doi:10.1109/JIOT.2018.2840129

Gao, P., Wang, M., Ghiocel, S. G., Chow, J. H., Fardanesh, B., and Stefopoulos, G. (2016). Missing Data Recovery by Exploiting Low-Dimensionality in Power System Synchrophasor Measurements. *IEEE Trans. Power Syst.* 31 (2), 1006–1013. doi:10.1109/TPWRS.2015.2413935

Khaledian, E., Pandey, S., Kundu, P., and Srivastava, A. K. (2021). Real-Time Synchrophasor Data Anomaly Detection and Classification Using Isolation Forest, KMeans, and LoOP. *IEEE Trans. Smart Grid* 12, 2378–2388. *IEEE Trans. Smart Grid* (Early Access)in press. doi:10.1109/TSG.2020.3046602

Khan, M., Li, M., Ashton, P., Taylor, G., and Liu, J. (2014). "Big Data Analytics on PMU Measurements," in 2014 11th International Conference on Fuzzy Systems and Knowledge Discovery (FSKD). Xiamen, China. August 19-21, 2014, 715–719.

Kumar, V. S., Wang, T., Aggour, K. S., Wang, P., Hart, P. J., and Yan, W. (2021). "Big Data Analysis of Massive PMU Datasets: A Data Platform Perspective," in 2021IEEE Power & Energy Society Innovative Smart Grid Technologies Conference (ISGT). Washington, DC, USA. February 17-20, 2021, 1–5.

Kummerow, A., Nicolai, S., and Bretschneider, P. (2018). "Ensemble Approach for Automated Extraction of Critical Events from Mixed Historical PMU Data Sets," in 2018 IEEE Power & Energy Society General Meeting (PESGM). Portland, OR. August 20-25, 2018, 1–5.

Lathiya, P., and Rani, R. (2016). "Improved CURE Clustering for Big Data Using Hadoop and Mapreduce," in 2016 International Conference on Inventive Computation Technologies (ICICT). Coimbatore, India. August 26-272016, 1–5.

Li, Q., Xu, Y., Ren, C., and Zhao, J. (2019). "A Hybrid Data-Driven Method for Online Power System Dynamic Security Assessment with Incomplete PMU Measurements," in 2019 IEEE Power & Energy Society General Meeting (PESGM). Atlanta, GA, USA. August 4-8, 2019, 1–5.

Liu, F. T., Ting, K. M., and Zhou, Z. (2008). "Isolation Forest," in 2008 Eighth IEEE International Conference on Data Mining, December 15-19, 2008. Pisa, Italy, 413–422.

Liu, J., Cao, Y., Li, Y., Guo, Y., and Deng, W. (2020). A Big Data Cleaning Method Based on Improved CLOF and Random Forest for Distribution Network. *CSEE J. Power Energ. Syst.* 1 –10. doi:10.17775/CSEEJPES.2020.04080

Ma, J., and Perkins, S. (2003). "Time-series novelty Detection Using One-Class Support Vector Machines," in Proceedings of the International Joint Conference on Neural Networks. PortlandUSA: OR. July 20-24, 2003, 1741–1745.

Mahapatra, K., Chaudhuri, N. R., and Kavasseri, R. (2016). "Bad Data Detection in PMU Measurements Using Principal Component Analysis," in 2016 North American Power Symposium (NAPS), September 18-20, 2016. Denver, CO, USA, 1–6.

Mahapatra, K., Chaudhuri, N. R., and Kavasseri, R. (2017). "Online Bad Data Outlier Detection in PMU Measurements Using PCA Feature-Driven ANN Classifier," in 2017 IEEE Power & Energy Society General Meeting (PESGM). July 16-20, 2017. Chicago, IL, USA, 1–5.

Manh, T. T., and Kim, J. (2011). "The Anomaly Detection by Using DBSCAN Clustering with Multiple Parameters," in 2011 International Conference on Information Science and Applications. April 26-29, 2011. Jeju, Korea (South), 1–5.

Menon, V. K., Variyar, V. S., Soman, K. P., Gopalakrishnan, E. A., Kottayil, S. K., Almas, M. S., et al. (2018). "A Spark™ Based Client for Synchrophasor Data Stream Processing," in 2018 International Conference and Utility Exhibition on Green Energy for Sustainable Development (ICUE). Phuket, Thailand. October 1-32018, 1–9.

Mrabet, Z. E., Selvaraj, D. F., and Ranganathan, P. (2019). "Adaptive Hoeffding Tree with Transfer Learning for Streaming Synchrophasor Data Sets," in 2019 IEEE International Conference on Big Data (Big Data), December 9-12, 2019. Los Angeles, CA, USA, 5697–5704.

Pokrajac, D., Lazarevic, A., and Jan, L. L. (2007). "Incremental Local Outlier Detection for Data Streams," in 2007 IEEE Symposium on Computational Intelligence and Data Mining. March 1-April 5.2007. Honolulu, HI, USA, 504–515.

Reif, M., Goldstein, M., Stahl, A., and Breuel, T. M. (2008). "Anomaly Detection by Combining Decision Trees and Parametric Densities," in 2008 19th International Conference on Pattern Recognition, December 8-11, 2008. Tampa, FL, USA, 1–4.

Ren, C., and Xu, Y. (2019). A Fully Data-Driven Method Based on Generative Adversarial Networks for Power System Dynamic Security Assessment with Missing Data. *IEEE Trans. Power Syst.* 34 (6), 5044–5052. doi:10.1109/TPWRS.2019.2922671

Ren, H., Hou, Z. J., Vyakaranam, B., Wang, H., and Etingov, P. (2020). Power System Event Classification and Localization Using a Convolutional Neural Network. *Front. Energ. Res.* 8, 607826–607837. doi:10.3389/fenrg.2020.607826

Wang, C., Cao, Y., Zhang, S., and Ling, T. (2021). A Reconstruction Method for Missing Data in Power System Measurement Based on LSGAN. *Front. Energ. Res.* 9, 651807–651820. doi:10.3389/fenrg.2021.651807

Wu, D., jiang, Z., Xie, X., Wei, X., Yu, W., and Li, R. (2020). LSTM Learning with Bayesian and Gaussian Processing for Anomaly Detection in Industrial IoT. *IEEE Trans. Ind. Inf.* 16 (8), 5244–5253. doi:10.1109/TII.2019.2952917

Yang, B., Yamazaki, J., Saito, N., Kokai, Y., and Xie, D. (2015). "Big Data Analytic Empowered Grid Applications — Is PMU a Big Data Issue?," in 2015 12th International Conference on the European Energy Market (EEM). May 19-22, 2015. Lisbon, Portugal, 1–4.

Yang, Z., Liu, H., Bi, T., Yang, Q., and Xue, A. (2019). "A PMU Data Recovery Method Based on Feature Component Extraction," in 2019 IEEE Power & Energy Society General Meeting (PESGM). August 4-8. Atlanta, GA, USA, 1–5.

Zhao, J., Qi, J., Huang, Z., Meliopoulos, A. P. S., Gomez-Exposito, A., Netto, M., et al. (2019). Power System Dynamic State Estimation: Motivations, Definitions, Methodologies, and Future Work. *IEEE Trans. Power Syst.* 34 (4), 3188–3198. doi:10.1109/TPWRS.2019.2894769

Zhou, M., Wang, Y., Srivastava, A. K., Wu, Y., and Banerjee, P. (2019). Ensemble-Based Algorithm for Synchrophasor Data Anomaly Detection. *IEEE Trans. Smart Grid* 10 (3), 2979–2988. doi:10.1109/TSG.2018.2816027

Conflict of Interest: Authors LS, XH, ML, RQ, CG, XM, and RD were employed by the company Electric Power Research Institute of Yunnan Power Grid Company Ltd.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Energy Research

ORIGINAL RESEARCH
published: 05 August 2021
doi: 10.3389/fenrg.2021.720391

Nonintrusive Monitoring for Electric Vehicles Based on Zero-Shot Learning

*Jingwei Hu, Rufei Ren, Jie Hu and Qiuye Sun**

College of Information Science and Engineering, Northeastern University, Shenyang, China

Monitoring the charging behavior of electric vehicle clusters will contribute to developing more effective energy management strategies for grid operators. A low implementation cost leads to a wide application prospect in nonintrusive monitoring for EVs. Aiming at the problem that traditional nonintrusive monitoring methods cannot identify unknown devices accurately due to the lack of classes, a nonintrusive monitoring method based on zero-shot learning (ZSL) is proposed in this article, one which can monitor the unknown types of EVs connected to charging piles. First, the charging characteristics of known EVs and unknown EVs are extracted by dictionary learning. Then EVs are classified by ZSL based on sparse coding. Furthermore, EVs are decomposed based on the proposed multimode factorial hidden Markov model (FHMM). Finally, the EV dataset of Pecan Street is used to verify the effectiveness and accuracy of the proposed method.

Keywords: nonintrusive monitoring, electric vehicles, zero-shot learning, factorial hidden markov model, long short-term memory

OPEN ACCESS

Edited by:
Peng Li,
Tianjin University, China

Reviewed by:
Pengfei Tu,
Nanyang Technological University, Singapore
Wei Hu,
Zhejiang University, China

****Correspondence:***
Qiuye Sun
sunqiuye@ise.neu.edu.cn

Specialty section:
This article was submitted to Smart Grids, a section of the journal Frontiers in Energy Research

Received: *04 June 2021*
Accepted: *02 July 2021*
Published: *05 August 2021*

Citation:
Hu J, Ren R, Hu J and Sun Q (2021) Nonintrusive Monitoring for Electric Vehicles Based on Zero-Shot Learning. Front. Energy Res. 9:720391. doi: 10.3389/fenrg.2021.720391

1 INTRODUCTION

With the continuous improvement of the penetration of renewable energy, the gradual decline in the electricity price has been making EVs more appealing to consumers (Liu et al., 2013). The promotion of EVs is regarded as one of the effective means to achieve energy conservation and emission reduction. When a large number of EVs are connected to the grid for charging and discharging, EVs are no longer just transport agents but also controllable loads and distributed energy sources in the energy system (Sun et al., 2019; Wang et al., 2021b). In this sense, EVs are distributed and mobile energy storage units.

However, the random and uncertain behavior generated by the EVs' charging demand will have many negative effects on the grid (Ahmadian et al., 2015; Mehta et al., 2018). These effects include increased peak power demand and overloads on feeders and transformers, especially in the distribution network (Li et al., 2019; Wang et al., 2020a). Therefore, it is necessary to monitor the charging and discharging behavior of EVs.

According to the EV charging environment, EV monitoring can be divided into household EV monitoring and parking lot EV monitoring (Rastogi et al., 2019). At present, the minimum charging and discharging power of EVs sold on the market is 3 KW, which is 5–10 times that of most household appliances. When the household charging pile is connected with EVs, the power fluctuation is obvious. For parking lots equipped with charging piles, most of them are planned in buildings, such as apartments, office buildings, and shopping malls. Due to the high power demand of EVs, it is significant to monitor such EV clusters:

1) For the parking lot with a large number of charging piles, its charging behavior has a certain regularity and synchronization, and such cluster charging behavior will generate a surge in the electricity demand (Li et al., 2021). In order to ensure intelligent charging of EVs, the

power grid/third-party assistance can monitor EV charging in a more intelligent way (Dickerman and Harrison, 2010; Rastogi et al., 2019);

2) Charging an EV takes longer than filling up an internal combustion engine vehicle. The real-time queuing information for EVs can be provided by monitoring the parking lot equipped with charging piles, including the number of EVs being charged and the estimated waiting time (de Weerdt et al., 2016; Goel et al., 2020);
3) In the long-term vision, one of the goals of the smart grid is to optimize the power service economy by establishing a two-way relationship between the power grid and EVs (Zeff, 2016). EVs can be regarded as energy storage. In the future, the advantages of short-term storage of EVs can be used to charge in the trough power demand and feed energy back to the grid in the peak power demand (Ahmadian et al., 2020; Wang et al., 2021a).

EV monitoring can be divided into intrusive monitoring and nonintrusive monitoring (NIM). Measurement data of each charging pile are required to be collected in intrusive EV monitoring, so as to realize real-time charging and discharging power monitoring of EVs. This intrusive monitoring method has a high cost of installation and maintenance due to the need to install a data acquisition device on each charging pile. The nonintrusive monitoring method was proposed by Hart in 1992. This method only needs the aggregated data measured at a single metering point to effectively realize the monitoring and identification of EVs.

In recent years, in the field of nonintrusive monitoring, many scholars have devoted themselves to improving the accuracy and applicability of NIM technology. Various methods have been used for NIM. An energy decomposition algorithm based on the adversarial network and the joint adaptation network is applied to NIM, which reduces the distribution gap of the feature space and the label space between the source domain and the target domain (Liu et al., 2021). A multitask NIM model based on the deep neural network is proposed, which can simultaneously analyze energy estimation and load state detection (Cimen et al., 2021). A hybrid event detection method is used for NIM for devices with long transients, high fluctuations, and/or near simultaneous action (Lu and Li, 2020). A multi-label classification method based on sparse representation classification is proposed, which can realize a fuzzy clustering algorithm inspired by NIM competitive-aggregation constrained by the entropy index through less training data (Singh and Majumdar, 2020). A low-complexity unsupervised NIM algorithm is proposed for the use of devices in families (Liu et al., 2019). In view of device feature representation in event-based NIM, Faustine et al. combined the adaptive weighted recursive graph block with the deep neural network architecture for device identification (Faustine et al., 2021). A convolutional neural network based on multi-scale features and context information is used to improve the accuracy of load decomposition (Chen et al., 2020). In order to improve the accuracy of new data decomposition, Hasan Rafiq trained the deep convolutional neural network model through data expansion (Rafiq et al., 2021). Taking advantage of the fact that the HMM can model multimode devices separately, a layered hidden Markov model (HHMM) is used for load decomposition of household appliances, one which can conduct nonintrusive monitoring of appliances with multiple modes and different power consumptions (Kong et al., 2018).

Recently, with the popularity of EVs, nonintrusive EV monitoring has gradually attracted the attention of scholars. On the basis of NIM, a training-free, nonintrusive load extraction algorithm was proposed based on boundary box fitting and load characteristics (Zhao et al., 2019), which can automatically identify the start time, end time, and power amplitude of charging events. Based on the low-frequency characteristics of the charging load mode, a charging load extraction method based on residential smart meter data was proposed to realize the nonintrusive extraction of the residential EV charging load mode (Xiang et al., 2021). Based on independent component analysis, an unsupervised EV charging load extraction method is proposed in the study by Munshi and Mohamed (2019). The proposed algorithm only requires the low-frequency active power measurement data. A nonintrusive identification method for EV charging curve extraction driven by a depth generation model is proposed in the study by Wang et al. (2020b). The proposed Markov model embedded in the presentation layer can solve the likelihood distribution overlap of learning.

However, there are many types of EVs and various battery types in EVs, and NIM needs to know the types and characteristics of monitored objects in advance, which greatly limits the accuracy of monitoring EVs.

Zero-shot learning (ZSL) provides an effective solution to the problem of class absence in nonintrusive EV monitoring. ZSL refers to the technology of using some known category data and the auxiliary information corresponding to the known category to train a certain model, so as to realize the classification and recognition of the data of the unknown category. A ZSL approach is proposed to simulate knowledge transfer between classes by learning visually consistent word vectors and tag embedding models (Demirel et al., 2019). The main idea is to project the vector space word vectors of attributes and classes into the visual space, so as to make the word representation of semantically related classes more close and, furthermore, use the proposed projection vector embedded in the model to identify the invisible classes. A transfer-sensing embedded projection method to solve multi-label ZSL learning was proposed in the study by Ye and Guo (2019). In this method, the label embedding vector is projected into a low-dimensional space to induce a better inter-label relationship, and the multi-label classifier with the largest boundary is learned *via* the projection label embedding. A ZSL classification method is proposed, which can automatically learn label embedding from input data in a semi-supervised large-profit learning framework, in the study by Li et al. (2015). A generation model is proposed, which simplifies the ZSL problem to a supervised classification task, in the study by Sariyildiz and Cinbis (2019). A ZSL method based on unsupervised domain adaptation was proposed in the study by Kodirov et al. (2015). In order to overcome the problem

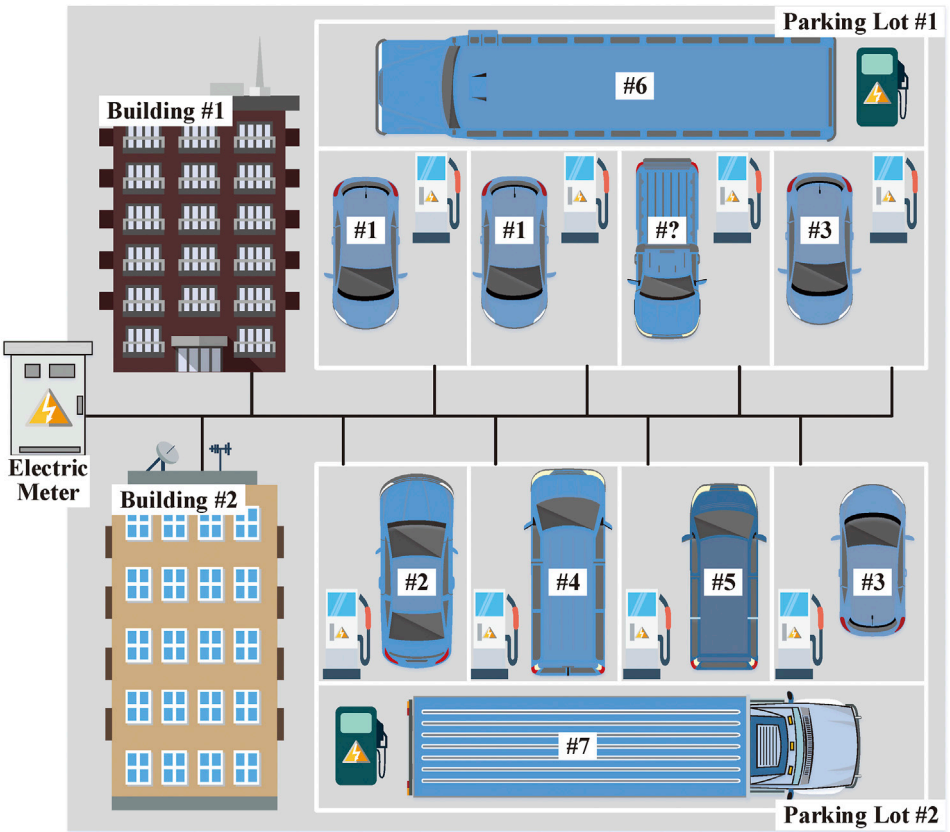

FIGURE 1 | Framework of nonintrusive monitoring for the electric vehicle cluster.

of the domain shift in the process of ZSL projection, the method regularized sparse coding to regularize the target domain projection.

In view of this, this study proposes a nonintrusive EV monitoring method based on the ZSL factor hidden Markov model. Firstly, the charge–discharge characteristics of known EVs and unknown EVs connected to the charging pile were extracted by dictionary learning. Furthermore, ZSL based on sparse coding is used to classify EVs. Finally, EVs are decomposed based on the proposed FHMM based on the bilateral long-term and short-term memory network (Bi-LSTM).

The rest of this article is organized as follows: **Section 2** describes the nonintrusive EV monitoring and extracts and classifies the charging and discharging status of EVs, **Section 3** proposes an FHMM-based EV decomposition method, **Section 4** discusses the proposed method of performing simulation and verification, and conclusions are drawn in **section 5**.

2 EVS' CHARGE–DISCHARGE STATUS EXTRACTION AND CLASSIFICATION

A framework of nonintrusive monitoring for EV clusters is given in **Figure 1**. The data monitored using the meter include the building load and the EV charging load. In the study by Munshi and Mohamed (2019), the existing EV charging loads are summarized into seven types, as shown in **Table 1**. However, as the market for EVs gradually expands, the charging power of EVs will also be varied. During nonintrusive monitoring of EVs' charge–discharge power, EVs of an unknown charge–discharge model are identified. In the parking lot of **Figure 1**, there are seven known-model EVs and one unknown-model EV.

TABLE 1 | Battery parameters of typical EVs.

Type	Power (kW)	Capacity (kWh)
Porsche Panamera	3	17.9
Nissan LEAF	3.3	24
e-Golf	3.6	35.8
BYD7009BEV2	7	76.9
ID. 3 Pure	7.2	45/58/75/77
Fiat 500e X	7.4	23.8/42
Tesla Model X	10	100

2.1 Definitions

The total power time series monitored using the meter is $x = \{x_1, x_2, \ldots, x_T\}$, where the EVs' charge–discharge power time series is $P = \{P_1, P_2, \ldots, P_T\}$. The parking lot has M charging piles, one charging pile supports only one EV access,

and the charge–discharge power time series of the $m-th$ charging pile is $p_m = \{p_{m,1}, p_{m,2}, \ldots, p_{m,T}\}$. $p_{m,t}$ represents the magnitude of the charging power at time period t for the $m-th$ charging pile. In other words, the total power curves of M charging piles represent the charge–discharge power time series of EVs in this region. In addition, a maximum of M EVs are charged and discharged at time period t.

Furthermore, in the time period t, $s_{m,t}$ represents the charge–discharge state of the charging pile, that is, if the charging pile is in the charging state, then $s_{m,t} = 1$; if the charging pile is in the stopping state, then $s_{m,t} = 0$; and if the EV is in the discharging state, then $s_{m,t} = -1$. It is worth noting that one charging pile may successively access multiple different types of EVs in the time series T. Although the charge–discharge state of the charging pile is a time series of $(0, 1, -1)$, the charge–discharge power may vary in size.

2.2 Description of the Problem

Suppose the time series of the EV charge–discharge power of the group N_{EV} is known, and the known dataset constituted by it is $D_{EV} = \{(p_{EV,i}, y_{EV,i})\}_{i=1}^{N_{EV}}$, where $p_{EV,i}$ represents the charge–discharge power time series of the EV of the group i and $y_{EV,i}$ represents the corresponding EV label. Each known class EV label comes from a collection Y_{EV} of known class labels, that is, $y_{EV,i} \in Y_{EV} = \{y_i\}_{i=1}^{Q_{EV}}$, where Q_{EV} is the number of known-type EVs. In addition, the time series of the charge–discharge power of the group N_X of EVs to be identified constitutes the unknown dataset $D_X = \{(p_{X,i}, y_{X,i})\}_{i=1}^{N_X}$. For each unknown power time series $p_{X,i}$, its label $y_{X,i}$ comes from the unknown class label set Y_X, that is, $y_{X,i} \in Y_X = \{y_i\}_{i=1}^{Q_X}$, where Q_X is the number of unknown-type EVs. The set Y_{EV} of known category labels and the set Y_X of unknown category labels constitute the complete set Y of categories, and the set of known category labels and the set of unknown category labels do not intersect with each other, that is, $Y_{EV} \cup Y_X = Y$, $Y_{EV} \cap Y_X = \varnothing$.

In most studies at home and abroad, the nonintrusive monitoring problem for low-frequency data is mostly regarded as a load decomposition problem, and the related technologies mainly have two subtasks: 1) classification and 2) reconstruction. First, the operation state of the device can be divided into known classes by classification, and second, the time series monitored using the meter can be reconstructed based on the classification results. Based on the purpose of nonintrusive monitoring, a nonintrusive EV recognition method based on zero-sample migration learning is proposed in this study. The classification model is trained by using the known dataset $D_{EV} = \{(p_{EV,i}, y_{EV,i})\}_{i=1}^{N_{EV}}$ of EVs and its auxiliary information, and the knowledge learned by the known dataset is effectively migrated to the unknown dataset, so as to realize the classification of the unknown EV dataset.

Considering the case of the unknown type of EVs accessing charging piles for charge–discharge, there is a mapping offset problem due to the poor generalization ability of the mapping model when classifying the operating state of EVs. It is shown that in the training process, the model maps the time series of known EVs' charge–discharge power into the known class label space in the semantic space, and due to the lack of the unknown dataset composed of the time series of EVs' charge–discharge power to be identified, the classification model will not map the unknown dataset into the unknown class label space at the time of testing, that is, there is a mapping offset, and the unknown EV category cannot be accurately identified.

2.3. Zero-Shot Classification Based on Sparse Coding

To solve the mapping offset problem of unknown-type EVs in the recognition process, this study converts the projection function learning problem into a sparse coding problem using the unsupervised domain adaptive model proposed in the study by Kodirov et al. (2015) as follows: each dimension of the semantic embedding space corresponds to the dictionary base vector, and the sparse code of each feature vector is its projection in the semantic embedding space. Regularity terms are introduced separately for the dictionary learning problem of the charge–discharge power time series of known-type EVs and unknown-type EVs. The known-type EVs' semantic dictionary learning problem can be expressed as follows:

$$L_{EV} = \min_{L_{EV}} \|P_{EV} - L_{EV}H_{EV}\|_F^2 + \lambda \|L_{EV}\|_2^2, s.t. \ \|l_i\|_2^2 \le 1, \quad (1)$$

where L_{EV} is the semantic dictionary of known-type EVs, H_{EV} is the semantic representation of known-type EVs, $\|\cdot\|_F$ is the Frobenius norm, and λ controls the strength of the regular term.

Unlike known-type EVs' semantic dictionary learning, in unknown-type EVs' semantic dictionary learning, both the unknown-type EV semantic dictionary L_X and the unknown-type EV semantic representation S_X are unknown. To overcome the domain offset problem during learning, that is, mapping offset, L_{EV} and H_X time is used to optimize L_X. Therefore, the unknown-type EVs' semantic dictionary learning problem can be expressed as follows:

$$\begin{aligned} \{L_X, H_X\} = \min_{L_X, S_X} & \|P_X - L_X H_X\|_F^2 + \lambda_1 \|L_X - L_{EV}\|_F^2 \\ & + \lambda_2 \sum_{i,j} \omega_{ij} \left\| s_i - q_j \right\|_2^2 + \lambda_3 \|L_X\|_1, \\ & s.t. \ \|l_i\|_2^2 \le 1, \end{aligned} \quad (2)$$

where $\|L_X\|_1 = \sum_{i=1}^{N_X} \|l_i\|_1$, q_j is the representation of $y_{X,j}$ in the semantic embedding space, and ω_{ij} is the probability that the input time series belongs to the label $y_{X,j}$, whose size can be estimated using the IAP model (Lampert et al., 2009). The first regular term $\|L_X - L_{EV}\|$ is used to limit the fitness of L_X vs. L_{EV}, and the second regular term $\|h_i - q_j\|$ is used to limit the similarity of the representation of unknown-type EVs' power time series in the semantic embedding space to the representation of unknown-type EV labels in the semantic embedding space. Based on **Eq. 3**, the method of alternating iteration is used to solve one, and H_X and L_X are fixed to solve another, as follows:

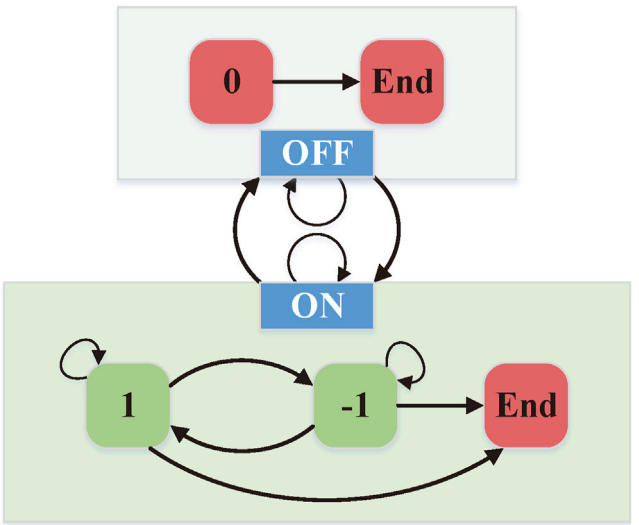

FIGURE 2 | Schematic diagram of EV charging–discharge state transfer.

$$\begin{aligned} L_X^* &= \arg\min_{L_X} \|P_X - L_X H_X\|_F^2 + \lambda_1 \|L_X - L_{EV}\|_F^2 \\ H_X^* &= \arg\min_{H_X} \|P_X - L_X H_X\|_F^2 + \lambda_2 \sum_{i,j} \omega_{ij} \left\| h_i - q_j \right\|_2^2 + \lambda_3 \|L_X\|_1 \end{aligned}. \quad (3)$$

3 NONINTRUSIVE HIDDEN MARKOV LOAD DECOMPOSITION

3.1 FHMM Model

The HMM can well describe the influence of the system equipment state change on the system output. The traditional HMM structure is given by, and its model θ can be expressed as follows:

$$\theta = (A, B, \pi), \quad (4)$$

where A is the state transition matrix, B is the observation matrix, and π is the initial state probability distribution, which is expressed as follows:

$$\begin{aligned} \pi &= \varphi(s_1 = i), \\ A &= [a_{ij}] = \varphi(s_t = j | s_{t-1} = i), \\ B &= \varphi(x_t | s_t = i) \sim N(\mu_i, \varepsilon_i), \end{aligned} \quad (5)$$

where $\varphi(x_t | s_t = i)$ denotes the output probability of the state $s_t = i$ to the observation matrix x_t, which obeys a normal distribution, μ_i is the mean vector, and ε_i is the covariance matrix of the observation matrix.

In the HMM-based nonintrusive monitoring problem of EVs, π and A determine the EVs' charge–discharge state sequence, and B determines the EVs' charge–discharge power time series $P_{1:T} = \{p_1, p_2, \ldots, p_T\}$. The nonintrusive monitoring of EVs based on the HMM can be divided into two stages: model parameter estimation and observation matrix decoding. In the model estimation stage, the model parameters are estimated by unsupervised learning or supervised learning, so that the probability of observed EVs' charge–discharge power time series under this model parameter reaches the largest, that is, $\theta^* = \arg\max \varphi(x|\theta)$; in the observation matrix decoding stage, the charge–discharge status and charging power of each charging pile are decomposed according to the optimal parameter θ^* estimated in the previous stage and the input EV charge–discharge power time series x_t.

The charge–discharge states of EVs can be represented by **Figure 2**. EVs' charge–discharge state can be divided into the OFF state and the ON state. The OFF state represents that EVs are in the standby state due to system scheduling or full battery power, which is equivalent to ending the charge–discharge behavior, that is, changing from OFF to End; the ON state represents that EVs are in the charge–discharge state, and there is a certain probability of ending the charge–discharge state. According to the EV charge–discharge state transfer rule, the charge–discharge state of the charging pile during the time series T can be expressed as $S_{1:T} = \{s_1, s_2, \ldots, s_T\}$, where $s_t = \{-1, 0, 1\}$.

According to the EV cluster monitoring framework depicted in **Figure 1**, the nonintrusive monitoring structure based on the FHMM is shown in **Figure 3**. Since the observable time series in this FHMM framework is the total power output containing the building and parking lot charging stakes, and not the EV charge–discharge power time series, $P_{1:T}$ needs to be extracted from $X_{1:T} = \{x_1, x_2, \ldots, x_T\}$.

3.2. Bi-LSTM Model

In the process of extracting $P = \{p_1, p_2, \ldots, p_T\}$, upon considering the state of charge (SOC) of the battery in EVs, it is not only determined by the current moment t and the future period $t+$ charge–discharge power but also by the charge–discharge power of the past period $t-$. Conventional LSTM performs forward transfer updating of the hidden layer state *via* one-way time series input when training, while full epoch data of $X_{1:T}$ are required when extracting EVs' charge–discharge time series. Based on this, in this study, $P_{1:T}$ is extracted using Bi-LSTM, and its structure is shown in **Figure 4**. The output expression of the bidirectional LSTM is as follows:

$$p_{\vartheta,t}^{join} = \left[p_{\vartheta,t}^{forw}, p_{\vartheta,t}^{back} \right], \vartheta \in [1, \upsilon], \quad (6)$$

where $p_{\vartheta,t}^{forw}$, $p_{\vartheta,t}^{back}$, and $p_{\vartheta,t}^{join}$ are the outputs of the forward LSTM neural network, the reverse LSTM neural network, and the Bi-LSTM in the hidden layer of layer ϑ at time t, respectively; υ is the number of hidden layers. The LSTM network used in this study contains six layers, in which the length of the input layer is the length of the time window t, the second layer is the convolution layer, which is used to extract features from the signal, the third and fourth layers are Bi-LSTM, the fifth layer is the convolution layer, and the sixth layer is the full connection layer. The whole network is

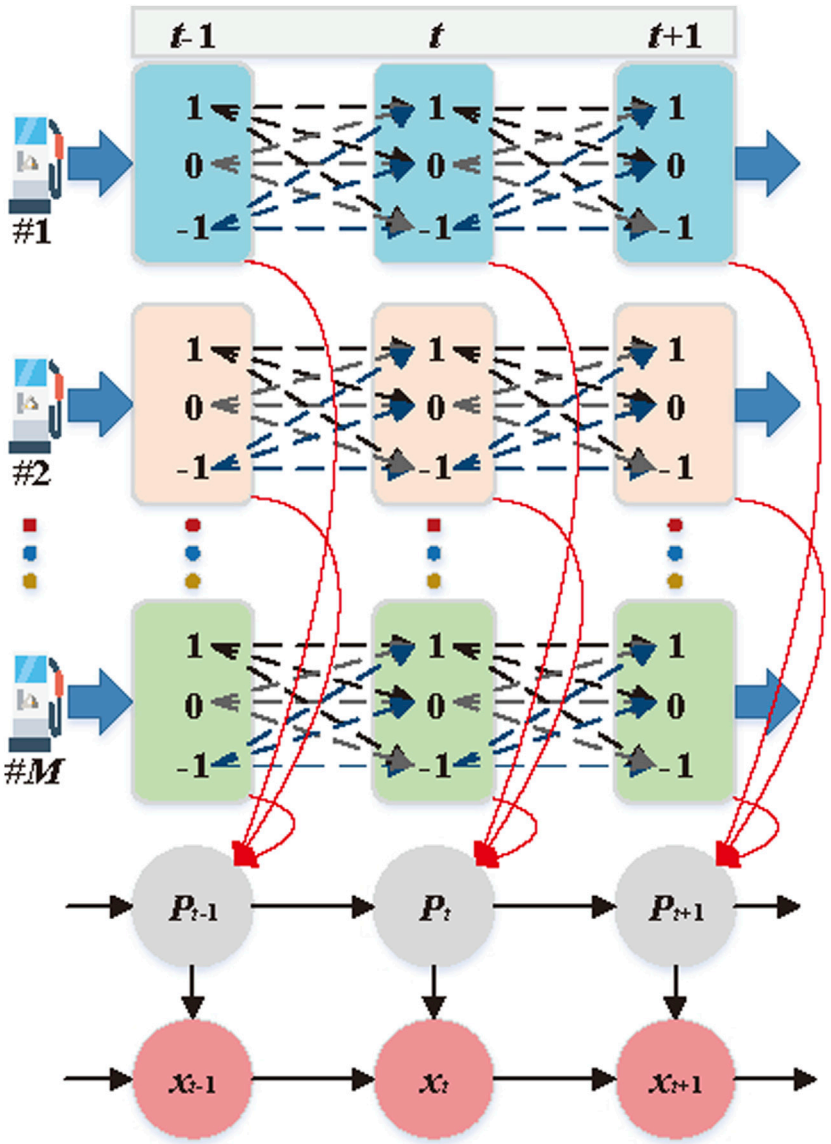

FIGURE 3 | Framework of nonintrusive monitoring for the EV cluster.

trained using the time forward and reverse bidirectional propagation method.

3.3 Nonintrusive EV Decomposition

In the framework shown in **Figure 3**, we note that although the charge–discharge time series of each charging pile are independent of each other, that is, M parallel Markov chains. Compared with the 3^M state combinations of HMM charging piles at time t, the decomposition of EVs based on the FHMM effectively reduces the complexity of the decomposition algorithm.

Considering that the known types of EVs are limited and the charge–discharge time series of the same type of EVs are similar, there may be multiple EVs of the same type for charge–discharge at time t. Based on the EVs' charge–discharge characteristics, this study further reduces the computational complexity of the EV decomposition model and classifies the same type of EV charge–discharge state at time t into a Markov chain. The improved HMM structure in **Figure 5** shows that there are $Q+1$ types of EVs, of which Q types of EVs are known, and $(Q+1)th$ is an unknown type of EV accessed by the charging pile. Here, the number of each type of EV accessed by the charging pile is assumed to be $\{d_1, d_2, \ldots, d_Q\}$. Therefore, the charge–discharge state of EVs of type $q-th$ at moment t can be expressed as follows:

$$s_t^{(q)} = \left\{1, 1-\frac{1}{d_q}, 1-\frac{2}{d_q}, \ldots, 0, -\frac{1}{d_q}, -\frac{2}{d_q}, \ldots, -1\right\} \quad (7)$$

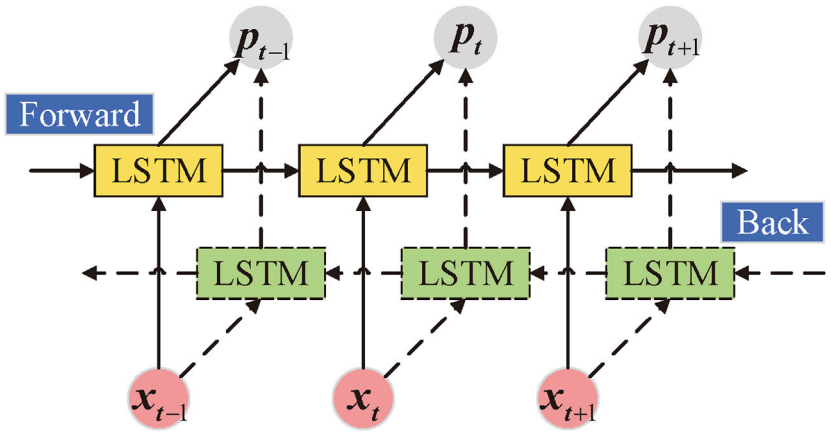

FIGURE 4 | Operating characteristics of the MT.

In the process of EV decomposition, considering that there are many unknown types of EVs connected to the charging pile and the charge–discharge power of unknown types of EVs is different, because the specific model of unknown types of EVs cannot be given, all unknown types of EVs' charge–discharge behavior is equivalent to a storage battery with unknown charge–discharge behavior in this study. When one unknown type of EV is connected to the charging pile, the hidden state is $\{1, 0, -1\}$, and when two unknown types of EVs are connected to the charging pile, the hidden state is $\left\{1, 1-\frac{1}{l_1}, 1-\frac{1}{l_2}, 1-\frac{l_1+l_2}{l_1 l_2}, 0, -\frac{1}{l_1}, -\frac{1}{l_2}, -\frac{l_1+l_2}{l_1 l_2}, -1\right\}$.

The nonintrusive FHMM model parameters based on clusters of EVs can be expressed as follows:

$$\begin{aligned} \pi &= \varphi\left(s_1^{(1)}, s_1^{(2)}, \ldots, s_1^{(Q+1)}\right), \\ A &= \varphi\left(s_t^{(1)}, s_t^{(2)}, \ldots, s_t^{(Q+1)} \middle| s_{t-1}^{(1)}, s_{t-1}^{(2)}, \ldots, s_{t-1}^{(Q+1)}\right), \\ B &= \varphi\left(p_t \middle| s_t^{(1)}, s_t^{(2)}, \ldots, s_t^{(Q+1)}\right) \sim N\left(\mu_i, \varepsilon_i\right). \end{aligned} \tag{8}$$

The FHMM model is composed of clusters of the same type of EV, each HMM chain represents the same type of EV, and the charge–discharge power and charge–discharge status are independent of each other between each type of EV. Therefore, the HMM model parameters constituted by each type of EV cluster are solved one by one, and thus obtain the total FHMM model parameters, such that the following occurs:

$$\theta^* = \arg\max_{\theta} \prod_{j=1}^{|P_{1:T}|} \sum_{s_t \in S} \varphi\left(P = p_t, S = s_t; \theta\right). \tag{9}$$

In this study, the EM (expectation maximization) algorithm is used to estimate the model parameters of **Eq. 9**. Two auxiliary variables, one forward variable $\alpha_t^{(q)}(i)$ and one backward variable $\beta_t^{(q)}(i)$, need to be used in the calculation. The forward variable $\alpha_t^{(q)}(i)$ represents the joint probability of an EV charge–discharge time series $p_1^{(n)}, \ldots, p_t^{(n)}$ of type $q-th$ at time t and the charge–discharge state i. Given the initial parameter θ_0, $\alpha_t^{(q)}(i)$ can be expressed as follows:

$$\alpha_t^{(q)}(i) = \varphi\left(s_t^{(q)} = i, p_1^{(n)}, \ldots, p_t^{(n)}, \theta_0\right). \tag{10}$$

Under the initial conditions, $\alpha_t^{(q)}(i)$ is expressed as follows:

$$\alpha_1^{(q)}(i) = \varphi\left(s_1^{(q)} = i\right)\varphi\left(p_1^{(n)} \middle| s_1^{(q)} = i\right). \tag{11}$$

Based on the $\alpha_1^{(q)}(i)$ forward recursion, the following occurs:

$$\alpha_{t+1}^{(q)}(j) = \sum_i \alpha_t^{(q)}(i) a_{ij}^{(q)} \varphi\left(p_{t+1}^{(n)} \middle| s_{t+1}^{(q)} = j\right). \tag{12}$$

The backward variable $\beta_t^{(q)}(i)$ represents the probability of observing an EV charge–discharge time series of type $q-th$ as $p_{t+1}^{(n)}, \ldots, p_T^{(n)}$ at time t and the charge–discharge state i. Given an initial parameter θ_0, $\beta_t^{(q)}(i)$ can be expressed as follows:

$$\beta_t^{(q)}(i) = \varphi\left(p_{t+1}^{(n)}, \ldots, p_T^{(n)} \middle| s_t^{(q)} = i, \theta_0\right). \tag{13}$$

Similarly, the backward variable $\beta_t^{(q)}(i)$ can also be calculated using the recursive formula as follows:

$$\beta_t^{(q)}(i) = \sum_j a_{ij}^{(q)} \varphi\left(p_{t+1}^{(n)} \middle| s_{t+1}^{(q)} = j\right)\beta_{t+1}^{(q)}(j), \tag{14}$$

where the initial value $\beta_T^{(q)}(i) = 1$.

Based on the two variables above, the FHMM decomposition process for a given initial parameter θ_0 and the observation sequence $P^{(q)} = [p_1^{(q)}, p_2^{(q)}, \ldots, p_T^{(q)}]$ can be divided into the following two steps:

1) E-Step: Compute the probability $\xi_t^{(q)}(i,j)$ of EVs of type $q-th$ shifting from state $s_t^{(q)} = i$ to state $s_{t+1}^{(q)} = j$ and the probability $\gamma_t^{(q)}(i)$ of presenting state $s_t^{(q)} = i$ at time t as follows:

$$\begin{aligned} \xi_t^{(q)}(i,j) &= \varphi\left(s_t^{(q)} = i, s_{t+1}^{(q)} = j \middle| P^{(q)}, \theta_0\right) \\ &= \frac{\alpha_t^{(q)}(i) a_{ij}^{(q)} \varphi\left(p_{t+1}^{(n)} \middle| s_{t+1}^{(q)} = j\right)\beta_{t+1}^{(q)}(j)}{\sum_{i=1}^{}\sum_{j=1}^{} \alpha_t^{(q)}(i) a_{ij}^{(q)} \varphi\left(p_{t+1}^{(n)} \middle| s_{t+1}^{(q)} = j\right)\beta_{t+1}^{(q)}(j)}, \end{aligned} \tag{15}$$

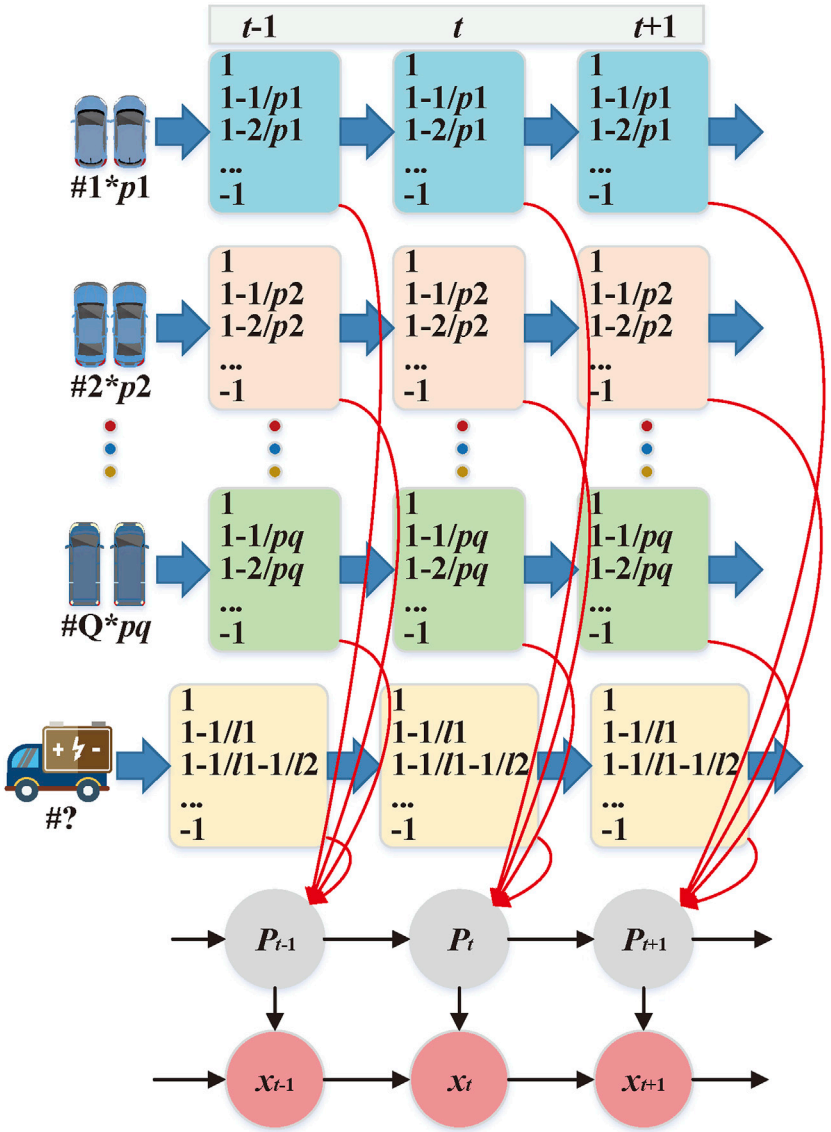

FIGURE 5 | Framework of nonintrusive monitoring for the EV cluster.

$$\gamma_t^{(q)}(i) = \varphi\left(s_t^{(q)} = i \middle| P^{(q)}, \theta_0\right) = \frac{\alpha_t^{(q)}(i)\beta_t^{(q)}(i)}{\sum_j \alpha_t^{(q)}(j)\beta_t^{(q)}(j)}. \tag{16}$$

2) M-Step: Recalculate model parameters for $q-th$ HMM chains as follows:

$$\hat{\pi}_i^{(q)} = \gamma_1^{(q)}(i), \tag{17}$$

$$\hat{a}_{ij}^{(q)} = \frac{\sum_{t=1}^{T}\xi_t^{(q)}(i,j)}{\sum_{t=1}^{T}\gamma_t^{(q)}(i,j)}, \tag{18}$$

$$\hat{\mu}_i^{(q)} = \frac{\sum_{t=1}^{T}\gamma_t^{(q)}(i)p_t^{(q)}}{\sum_{t=1}^{T}\gamma_t^{(q)}(i)}, \tag{19}$$

$$\hat{\varepsilon}_i^{(q)} = \frac{\sum_{t=1}^{T}\gamma_t^{(q)}(i)\left(p_t^{(q)} - \hat{\mu}_i^{(q)}\right)\left(p_t^{(q)} - \hat{\mu}_i^{(q)}\right)^{\mathrm{T}}\sum_{t=1}^{T}}{\sum_{t=1}^{T}\gamma_t^{(q)}(i)}. \tag{20}$$

3) Forward variables $\alpha_1^{(q)}(i)$ and backward variables $\beta_t^{(q)}(i)$, $\xi_t^{(q)}(i,j)$, and $\gamma_t^{(q)}(i)$ are iteratively calculated according to the new parameter cycle until convergence.

After all the parameters of the FHMM model are obtained, the hidden state can be decoded by applying the Viterbi algorithm to the summarized power consumption sequence. In the Viterbi algorithm, the variable $\delta_t(i)$ is introduced. We define $\delta_t(i)$ as the probability maximum of all states

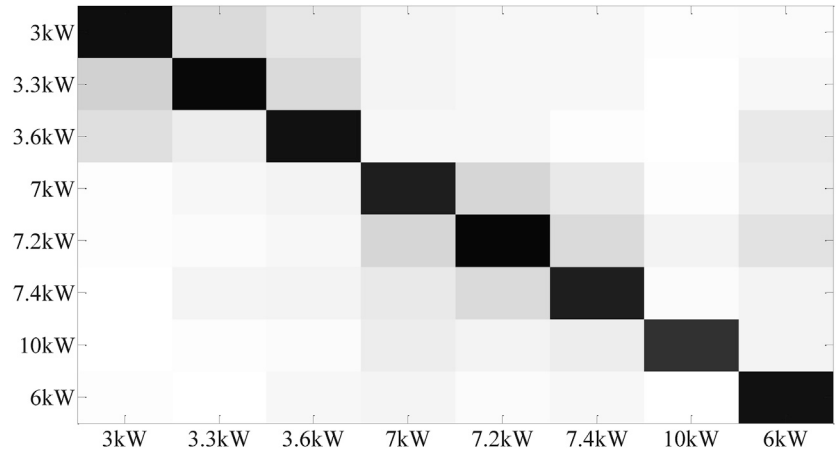

FIGURE 6 | Confusion matrices on different types of EVs.

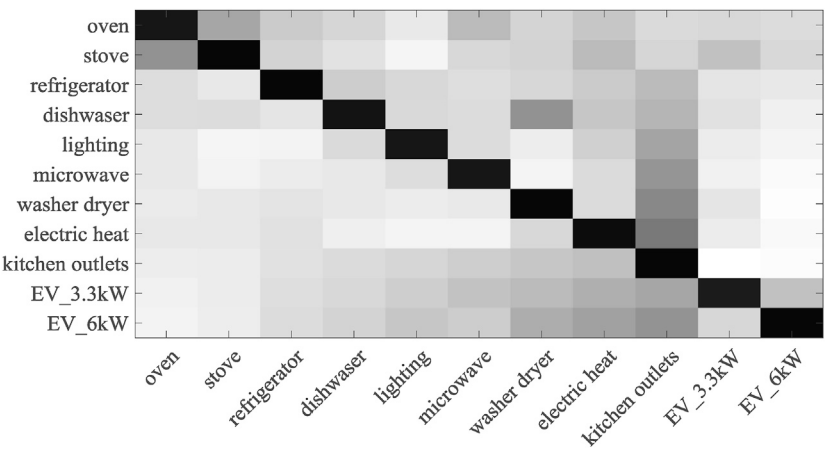

FIGURE 7 | Confusion matrices on different appliances.

$\{s_1, s_1, \ldots, s_t\}$ to the observation sequence $\{p_1, p_2, \ldots, p_T\}$ at the moment t as follows:

$$\delta_{t+1}(i) = \max_{1 \le j \le N} \left\{ \delta_t(j) a_{ji} \right\} \varphi\left(p_{t+1} \middle| s_{t+1} = i\right), \tag{21}$$

where N is the number of hidden layer states of EVs, $t = 1, 2, \ldots, T$.

The nonintrusive monitoring step of EVs based on the Viterbi algorithm is as follows:

1) input $\theta_0 = (\hat{\pi}_i, \hat{a}_{ij}, \hat{\mu}_i, \hat{\varepsilon}_i)$ and EV charge–discharge time series $P_{1:T} = \{p_1, p_2, \ldots, p_T\}$;
2) initialize $\delta_1(i) = \pi_i \varphi\left(p_1 \middle| s_1 = i\right)$;
3) recurrent $\delta_t(i) = \max_{1 \le j \le N} \{\delta_{t-1}(j) a_{ji}\} \varphi\left(p_t \middle| s_t = i\right)$; and
4) optimal path backtracking for $t = T-1, T-2, \ldots, 1$, $s_t(i) = \arg\max_{1 \le j \le N} \{\delta_t(j)\}$.

4 EXPERIMENT AND RESULT

4.1 Dataset Description

In order to verify that the method proposed in this study can classify unknown types of EVs effectively and, furthermore, monitor the charging behavior of EVs accurately, the processed EV dataset of Pecan Street is used for training and testing. Specifically, the dataset consists of 16,000 sets of EV charging data belonging to eight types. In the ZSL process, seven types of known EV data are used for training and one type of unknown EV data is used for testing, among which power levels of EVs are used as EV labels, that is, $Y_{EV} = \{3, 3.3, \ldots, 10\}$. The data sampling interval is 1 min. Charging services for EVs are provided by 120 primary/secondary charging piles.

TABLE 2 | Classification effects of different combinations of EV types.

Combination of EV types (kW)	Performance
3/7/10	−11.32/−12.30/−11.51
{3, 7}/{3, 10}/{7, 10}	−5.41/−6.37/−6.83
{3, 3.3, 3.6}	−3.44
{7, 7.2, 7.4}	−3.98
{3, 7, 10}	0.6
{3, 3.3, 3.6, 7}	1.21
{3, 7, 7.2, 7.4}	1.28
{3, 3.3, 3.6, 10}	1.37
{3, 7, 7.2, 7.4}	1.42
{3, 3.3, 7, 7.2}	1.48
{3, 3.6, 7, 10}	1.55

4.2 Zero-Shot Classification

In order to ensure that the method proposed in this study can accurately identify the access of unknown types of EVs in the EV cluster, this study selects seven kinds of EV charging data commonly available in the market for training and uses 6-kW EVs for testing. The test results are shown in **Figure 6**. It can be seen that zero-sample classification based on sparse coding can well identify the access of different types of EVs and has certain classification functions for unknown types of EVs. It is worth noting that the EV classification effect with a small power level difference is weaker than the EV classification effect with a large power level difference. Therefore, the classification effect of the {3kW, 3.3kW, 3.6kW} and {7kW, 7.2kW, 7.4kW} sets in **Figure 6** is not very obvious.

Considering the classification problem between EVs and other devices, this study took REDD centralized electric appliance equipment and 3.3-kW type EVs as the training set to test the EVs of {6kW}. The classification effect is shown in **Figure 7**.

In order to verify the recognition effect of different EV type combination training sets on unknown EV test sets, this study selects typical types of EV combinations for testing. The test sets include from one type of EV to four types of EV combinations. The recognition effect is shown in **Table 2**. It can be seen from **Table 2** that the more EV types there are, the better the recognition effect will be. In the case of a certain number of EV types, the more dispersed the EV power level is, the better the recognition effect will be.

To further verify the correctness of the conclusions above, this article studies the influence of the sample size of each EV type

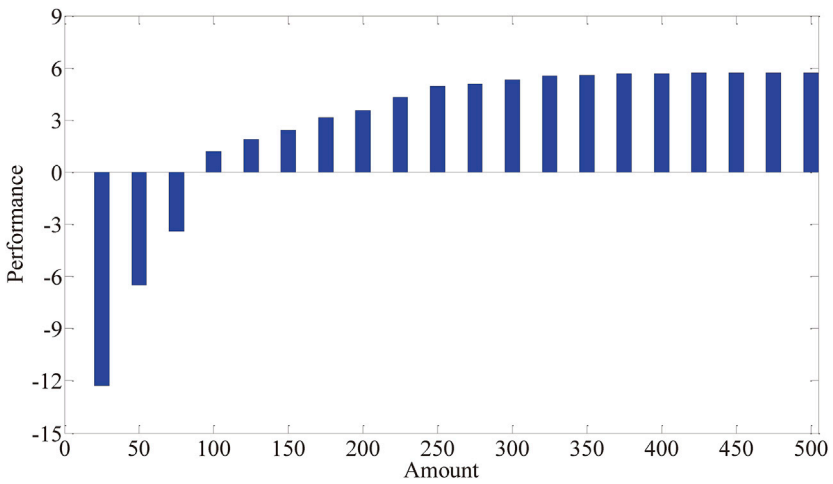

FIGURE 8 | Influence of the EV sample size on the recognition effect.

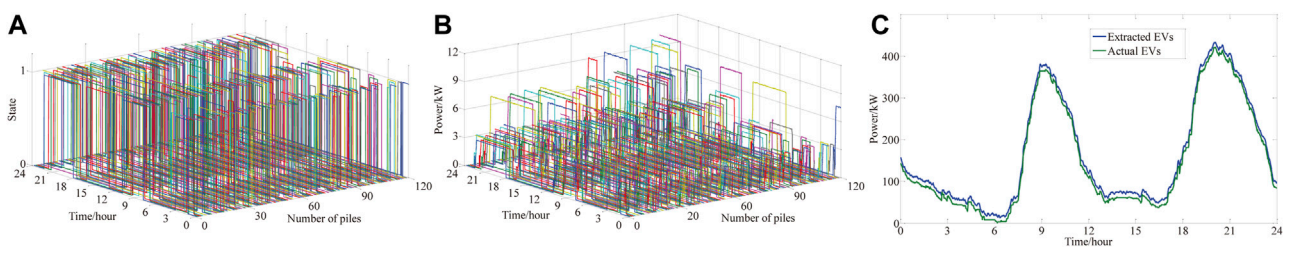

FIGURE 9 | Charging load decomposition effect of the traditional FHMM. **(A)** Charging state of 120 charging piles. **(B)** Charging power of charging piles. **(C)** Total charging power of EVs decomposed based on the FHMM.

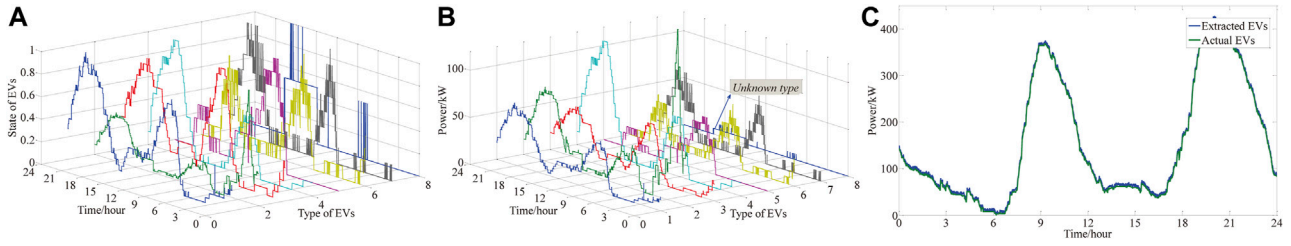

FIGURE 10 | Charging load decomposition effect of the traditional FHMM. **(A)** Charging states of EVs. **(B)** Charging power of each type of EV. **(C)** Total charging power of EVs decomposed based on the proposed method.

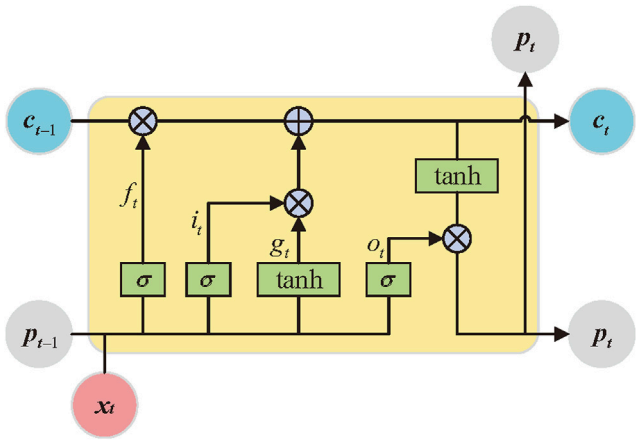

FIGURE 11 | Structure of LSTM.

training set on the recognition effect. The results are shown in **Figure 8**. It can be seen that the greater the number of samples in the training set, the more obvious the effect. At the same time, the performance tends to be stable when the sample size of the training set reaches 300.

4.3 EVs' Nonintrusive Decomposition

In order to verify the nonintrusive monitoring method of EVs proposed in this study, the decomposition effects of the traditional FHMM and the proposed method were compared. **Figure 9** shows the nonintrusive identification effect of the traditional FHMM, in which **Figure 9A** shows the charging state of 120 charging piles, **Figure 9B** shows the charging power of charging piles obtained based on LSTM, and **Figure 9C** shows the total charging power of EVs decomposed based on the FHMM.

The nonintrusive decomposition effect of EVs based on the method proposed in this study is shown in **Figure 10**. The charging states of EVs connected by charging piles is shown in **Figure 10A**, where the first 7 EVs are the charging status of known EVs, and the eighth is the charging status of unknown EVs.

Based on the decomposed EV charging state, the charging power of each type of EV can be obtained using Bi-LSTM. As shown in **Figure 10B**, the total power curve is synthesized according to the decomposed EV charging power. By comparing with **Figure 9C**, it can be seen that the nonintrusive decomposition of EVs proposed in this study has a higher precision.

5 CONCLUSION

In this study, the FHMM framework is used to develop a nonintrusive monitoring method based on ZSL. The charge–discharge characteristics of known EVs and unknown EVs connected to the charging pile were extracted by dictionary learning, and furthermore, the ZSL based on sparse coding was used to classify EVs, which could effectively identify the unknown EVs. In the validation process, the processed Pecan Street EV dataset was used for training and testing. Compared with the traditional FHMM, the nonintrusive decomposition of EVs proposed in this study has a higher accuracy.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

AUTHOR CONTRIBUTIONS

JH was responsible for the specific work of this article. JH and RR carried out some of the calculation work. QS guided the work of this article.

ACKNOWLEDGMENTS

The authors acknowledge the funding of the National Key Research and Development Program of China (2018YFA0702200) and the National Natural Science Foundation of China (U20A20190 and 62073065).

REFERENCES

Ahmadian, A., Mohammadi-Ivatloo, B., and Elkamel, A. (2020). A Review on Plug-In Electric Vehicles: Introduction, Current Status, and Load Modeling Techniques. *J. Mod. Power Syst. Clean Energ.* 8, 412–425. doi:10.35833/MPCE.2018.000802

Ahmadian, A., Sedghi, M., and Aliakbar-Golkar, M. (2015). "Stochastic Modeling of Plug-In Electric Vehicles Load Demand in Residential Grids Considering Nonlinear Battery Charge Characteristic," in 2015 20th Conference on Electrical Power Distribution Networks Conference (EPDC), Zahedan, Iran, April 28–29, 2015, 22–26. doi:10.1109/EPDC.2015.7330467

Chen, K., Zhang, Y., Wang, Q., Hu, J., Fan, H., and He, J. (2020). Scale- and Context-Aware Convolutional Non-intrusive Load Monitoring. *IEEE Trans. Power Syst.* 35, 2362–2373. doi:10.1109/TPWRS.2019.2953225

Cimen, H., Cetinkaya, N., Vasquez, J. C., and Guerrero, J. M. (2021). A Microgrid Energy Management System Based on Non-intrusive Load Monitoring via Multitask Learning. *IEEE Trans. Smart Grid* 12, 977–987. doi:10.1109/TSG.2020.3027491

de Weerdt, M. M., Stein, S., Gerding, E. H., Robu, V., and Jennings, N. R. (2016). Intention-aware Routing of Electric Vehicles. *IEEE Trans. Intell. Transport. Syst.* 17, 1472–1482. doi:10.1109/TITS.2015.2506900

Demirel, B., Cinbis, R. G., and Ikizler-Cinbis, N. (2019). "Learning Visually Consistent Label Embeddings for Zero-Shot Learning," in 2019 IEEE International Conference on Image Processing (ICIP), Taipei, China, September 22–25, 2019, 3656–3660. doi:10.1109/ICIP.2019.8803458

Dickerman, L., and Harrison, J. (2010). A New Car, a New Grid. *IEEE Power Energ. Mag.* 8, 55–61. doi:10.1109/MPE.2009.935553

Faustine, A., Pereira, L., and Klemenjak, C. (2021). Adaptive Weighted Recurrence Graphs for Appliance Recognition in Non-intrusive Load Monitoring. *IEEE Trans. Smart Grid* 12, 398–406. doi:10.1109/TSG.2020.3010621

Goel, S., Kumar, R., Kumar, A., and Malhotra, R. (2020). Smart Station Search Assistance for Electric Vehicle-A Step toward Smart City. *IEEE Consumer Electron. Mag.* 9, 27–33. doi:10.1109/MCE.2020.2985656

Kodirov, E., Xiang, T., Fu, Z., and Gong, S. (2015). "Unsupervised Domain Adaptation for Zero-Shot Learning," in 2015 IEEE International Conference on Computer Vision (ICCV), Santiago, Chile, December 7–13, 2015, 2452–2460. doi:10.1109/ICCV.2015.282

Kong, W., Dong, Z. Y., Hill, D. J., Ma, J., Zhao, J. H., and Luo, F. J. (2018). A Hierarchical Hidden Markov Model Framework for home Appliance Modeling. *IEEE Trans. Smart Grid* 9, 3079–3090. doi:10.1109/TSG.2016.2626389

Lampert, C. H., Nickisch, H., and Harmeling, S. (2009). "Learning to Detect Unseen Object Classes by Between-Class Attribute Transfer," in 2009 IEEE Conference on Computer Vision and Pattern Recognition, Miami, Florida, June 20–25, 2009, 951–958. doi:10.1109/CVPR.2009.5206594

Li, X., Guo, Y., and Schuurmans, D. (2015). "Semi-supervised Zero-Shot Classification with Label Representation Learning," in 2015 IEEE International Conference on Computer Vision (ICCV), Santiago, Chile, December 7–13, 2015, 4211–4219. doi:10.1109/ICCV.2015.479

Li, Y., Gao, W., Yan, W., Huang, S., Wang, R., Gevorgian, V., et al. (2021). Data-driven Optimal Control Strategy for Virtual Synchronous Generator via Deep Reinforcement Learning Approach. *J. Mod. Power Syst. Clean Energ.*, 1–11. doi:10.35833/MPCE.2020.000267

Li, Y., Zhang, H., Liang, X., and Huang, B. (2019). Event-triggered-based Distributed Cooperative Energy Management for Multienergy Systems. *IEEE Trans. Ind. Inf.* 15, 2008–2022. doi:10.1109/TII.2018.2862436

Liu, C., Chau, K. T., Wu, D., and Gao, S. (2013). Opportunities and Challenges of vehicle-to-home, Vehicle-To-Vehicle, and Vehicle-To-Grid Technologies. *Proc. IEEE* 101, 2409–2427. doi:10.1109/JPROC.2013.2271951

Liu, Q., Kamoto, K. M., Liu, X., Sun, M., and Linge, N. (2019). Low-complexity Non-intrusive Load Monitoring Using Unsupervised Learning and Generalized Appliance Models. *IEEE Trans. Consumer Electron.* 65, 28–37. doi:10.1109/TCE.2019.2891160

Liu, Y., Zhong, L., Qiu, J., Lu, J., and Wang, W. (2021). Unsupervised Domain Adaptation for Non-intrusive Load Monitoring via Adversarial and Joint Adaptation Network. *IEEE Trans. Ind. Inf.* 1, 1. doi:10.1109/TII.2021.3065934

Lu, M., and Li, Z. (2020). A Hybrid Event Detection Approach for Non-intrusive Load Monitoring. *IEEE Trans. Smart Grid* 11, 528–540. doi:10.1109/TSG.2019.2924862

Mehta, R., Srinivasan, D., Khambadkone, A. M., Yang, J., and Trivedi, A. (2018). Smart Charging Strategies for Optimal Integration of Plug-In Electric Vehicles within Existing Distribution System Infrastructure. *IEEE Trans. Smart Grid* 9, 299–312. doi:10.1109/TSG.2016.2550559

Munshi, A. A., and Mohamed, Y. A.-R. I. (2019). Unsupervised Nonintrusive Extraction of Electrical Vehicle Charging Load Patterns. *IEEE Trans. Ind. Inf.* 15, 266–279. doi:10.1109/TII.2018.2806936

Rafiq, H., Shi, X., Zhang, H., Li, H., Ochani, M. K., and Shah, A. A. (2021). Generalizability Improvement of Deep Learning-Based Non-Intrusive Load Monitoring System Using Data Augmentation. *IEEE Trans. Smart Grid* 12, 3265–3277. doi:10.1109/TSG.2021.3082622

Rastogi, S. K., Sankar, A., Manglik, K., Mishra, S. K., and Mohanty, S. P. (2019). Toward the Vision of All-Electric Vehicles in a Decade [energy and Security]. *IEEE Consumer Electron. Mag.* 8, 103–107. doi:10.1109/MCE.2018.2880848

Rui, W., Qiuye, S., Dazhong, M., and Xuguang, H. (2020a). Line Impedance Cooperative Stability Region Identification Method for Grid-Tied Inverters under Weak Grids. *IEEE Trans. Smart Grid* 11, 2856–2866. doi:10.1109/TSG.2020.2970174

Sariyildiz, M. B., and Cinbis, R. G. (2019). "Gradient Matching Generative Networks for Zero-Shot Learning," in 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Los Angeles, California, June 16–20, 2015, 2163–2173. doi:10.1109/CVPR.2019.00227

Singh, S., and Majumdar, A. (2020). Non-intrusive Load Monitoring via Multi-Label Sparse Representation-Based Classification. *IEEE Trans. Smart Grid* 11, 1799–1801. doi:10.1109/TSG.2019.2938090

Sun, Q., Zhang, N., You, S., and Wang, J. (2019). The Dual Control with Consideration of Security Operation and Economic Efficiency for Energy Hub. *IEEE Trans. Smart Grid* 10, 5930–5941. doi:10.1109/TSG.2019.2893285

Wang, R., Sun, Q., Hu, W., Li, Y., Ma, D., and Wang, P. (2021a). Soc-based Droop Coefficients Stability Region Analysis of the Battery for Stand-Alone Supply Systems with Constant Power Loads. *IEEE Trans. Power Electron.* 36, 7866–7879. doi:10.1109/TPEL.2021.3049241

Wang, R., Sun, Q., Tu, P., Xiao, J., Gui, Y., and Wang, P. (2021b). Reduced-order Aggregate Model for Large-Scale Converters with Inhomogeneous Initial Conditions in Dc Microgrids. *IEEE Trans. Energ. Convers.* 1, 1. doi:10.1109/TEC.2021.3050434

Wang, S., Du, L., Ye, J., and Zhao, D. (2020b). A Deep Generative Model for Non-intrusive Identification of Ev Charging Profiles. *IEEE Trans. Smart Grid* 11, 4916–4927. doi:10.1109/TSG.2020.2998080

Xiang, Y., Wang, Y., Xia, S., and Teng, F. (2021). Charging Load Pattern Extraction for Residential Electric Vehicles: A Training-free Non-intrusive Method. *IEEE Trans. Ind. Inf.* 1, 1. doi:10.1109/TII.2021.3060450

Ye, M., and Guo, Y. (2019). Multi-label Zero-Shot Learning with Transfer-Aware Label Embedding Projection. In 2019 IEEE International Conference on Image Processing (ICIP), Taipei, China, September 22–25, 2019, 3671–3675. doi:10.1109/ICIP.2019.8803720

Zeff, S. (2016). My Electric Journey with a Nissan Leaf: A Classic Early-Adopter Experience. *IEEE Consumer Electron. Mag.* 5, 79–80. doi:10.1109/MCE.2016.2556839

Zhao, H., Yan, X., and Ma, L. (2019). Training-free Non-intrusive Load Extracting of Residential Electric Vehicle Charging Loads. *IEEE Access* 7, 117044–117053. doi:10.1109/ACCESS.2019.2936589

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Energy Research

ORIGINAL RESEARCH
published: 25 August 2021
doi: 10.3389/fenrg.2021.732997

Design of Decentralized Adaptive Sliding Mode Controller for the Islanded AC Microgrid With Ring Topology

He Jiang[1,2]*, *Mofan Wei*[1,2], *Yan Zhao*[1,2] and *Ji Han*[1,2]

[1]School of Renewable Energy, Shenyang Institute of Engineering, Shenyang, China, [2]Key Laboratory of Regional Multi-energy System Integration and Control of Liaoning Province, Shenyang, China

Sliding mode control can restrain the perturbations generated from the intermittence of the renewable energy generation and the randomness of local loads when microgrids are operating in islanded mode. However, the microgrid consists of several subsystems and the interactions among them will cause the chattering problems under the overall sliding mode control. In this paper, the chattering restraint issues for voltage control of the islanded microgrid with a ring topology structure are investigated based on the decentralized adaptive sliding mode control strategies. Firstly, we construct a tracking error system with interconnections considering the power transmission among subsystems and nominal values of system states. Secondly, we design linear matrix inequalities (LMIs) according to the H_∞ attenuation performance of the system external disturbances. Then, the tracking error performance and the control precision are guaranteed via the asymptotic stabilities of integral sliding mode surfaces. Adaptive laws are utilized to address the chattering problems of the sliding mode control. Finally, simulation results verify the effectiveness of the proposed decentralized control methods.

Keywords: islanded microgrid, chattering restraint, voltage control, adaptive control, sliding mode control

OPEN ACCESS

Edited by:
Qiuye Sun,
Northeastern University, China

Reviewed by:
Wei Hu,
Zhejiang University, China
Jingwei Hu,
Northeastern University, China

***Correspondence:**
He Jiang
jianghescholar@163.com

Specialty section:
This article was submitted to Smart Grids, a section of the journal Frontiers in Energy Research

Received: *29 June 2021*
Accepted: *09 August 2021*
Published: *25 August 2021*

Citation:
Jiang H, Wei M, Zhao Y and Han J (2021) Design of Decentralized Adaptive Sliding Mode Controller for the Islanded AC Microgrid With Ring Topology. Front. Energy Res. 9:732997. doi: 10.3389/fenrg.2021.732997

INTRODUCTION

Recently, there are abundant distributed generating devices permeating in the modern electric power systems for achieving the environment protection and the effective and flexible control of grids. In order to ensure the extensiveness and security of the power supply, microgrids have been the main form to transmit electricity to local loads for remote regions, which can operate in islanded mode or grid-connected mode (Mahmoud et al., 2014). Actually, an AC islanded microgrid consisting of distributed generation units (DGus) and energy storage devices can supply power to local loads steadily in low voltage magnitude (Kabalan et al., 2017). Because the microgrid contains numerous power electronic facilities, such as voltage source converters (VSC) et al., it is lack of immense inertia provided by rotating devices comparing with conventional grids (Zou et al., 2019). Furthermore, the renewable generation devices are usually affected by weather conditions and the power generated from them is usually intermittent and uncertain, so it is more complicated to realize the stable control of the multi-area microgrid voltage when it is in islanded mode (Zhou et al., 2021). At present, there are different control strategies to solve the voltage control problems of multi-area microgrids and optimize the control performance in islanded mode in order to improve the reliability and effectiveness of the power supply (Divshali et al., 2012; Sahoo et al., 2018).

The conventional control methods for multi-area microgrids operating in islanded mode demonstrate several disadvantages. The close-loop proportional-integral-derivative (PID) voltage control strategy can not estimate the errors between state variables and the nominal sinusoidal voltages accurately. Additionally, this control strategy presents bad control performance for restraining the inner parameter perturbations, such as frequency fluctuations (Vandoorn et al., 2013; Chen et al., 2015; Sefa et al., 2015). Zeb et al. (2019) combined the PID control method with fuzzy principles and designed the proportional resonant harmonic compensator as a current controller. Moreover, a phase lock loop (PLL) was designed to promote the speed of the system dynamic response. A comparison between fuzzy sliding mode control (FSMC) and fuzzy PID control illustrated that the dynamic response speed was lower and the tracking error performance was less precise via the fuzzy PID method. Considering that microgrids are sensitive about the system parameter variations, so the droop control technology is proposed to improve the robustness of microgrids via simulating the droop relationships among different electrical parameters (Avelar et al., 2012; Beerten and Belmans, 2013; Eren et al., 2015; Wang et al., 2019; Wang et al., 2021). Mi et al. (2019) modified traditional linear droop control strategies and utilized nonlinear droop relationships to describe the interactions between reactive power and voltages. The T-S fuzzy theory was applied to approximate the nonlinear model accurately and coordinate power among each DGu. Nevertheless, there were also errors between stable values and nominal values of voltages. Recently, sliding mode control (SMC) strategies are extensively applied in the stability control of microgrids for the superior asymptotic stability and robustness against parameter uncertainties (Hu et al., 2010; Karimi et al., 2010; Liu et al., 2017). An integration model of microgrids with complex meshed topology structures and several DGus was constructed to achieve the power sharing and voltage robust control (Cucuzzella et al., 2017; Wang et al., 2020). But the integration model could not represent actual interaction effects in different subsystems and the chattering was serious. Mi et al. (2020) proposed an adaptive sliding mode control strategy based on the sliding mode observer for wind-diesel power systems. The microgrid bus voltage showed remarkable stability via regulating the reactive power in terms of this method. Contrarily, the disturbance observer and adaptive algorithm brought in numerous parameters and promoted the complexity of the control system. To figure out the problem of harmonic disturbance in microgrids, Esparza et al. (2017) proposed a comprehensive control strategy to restrain the harmonic currents generated from DGus in AC microgrids. As shown in the simulation results, this strategy could cause the chattering phenomenon inherently.

Motivated by the aforementioned discussions, for the multi-area microgrid with a ring topology, the decentralized voltage control model will represent more appropriate relationships among the parameters in each local subsystem comparing with the integrated one. In addition, the adaptive sliding mode control (ASMC) strategy, which will be designed according to H_∞ attenuation performance of each subsystem, can ensure the robustness of the interconnected systems against mismatched uncertainties and external perturbations.

The main contributions of this paper can be summarized as follows:

1) The established multi-area microgrid model can depict the interactions among subsystems appropriately;
2) The reliability of solutions and the attenuation performance of external disturbances can be ensured based on the linear matrix inequalities (LMIs);
3) The proposed decentralized ASMC can restrain the chattering of the microgrid.

The rest of research includes four sections. Section *Dynamical Models of Multi-Area Interconnected Microgrids* constructs state functions with interconnections representing the topology structure of microgrid systems and defines tracking error models based on the nominal values of the state variables. Section *Proposed Decentralized Adaptive Sliding Mode Voltage Controller* introduces the designed decentralized voltage controllers in terms of the proposed ASMC theory. Section *Simulation Results* provides the simulation results and Section *Conclusion* illustrates the conclusion.

DYNAMICAL MODELS OF MULTI-AREA INTERCONNECTED MICROGRIDS

In order to explain the power transmission among the multi-area microgrid, the electrical three-phase diagram of the ring topology system composed four DGus is shown in **Figure 1**. The researched microgrid consists of local loads, power transmission lines and DGus. Because of various energy storage components in renewable generation systems, the DGus could be represented as DC voltage sources. The DGus connect with the points of common coupling (PCC) via VSCs and filters and provide power to local loads. PCCs can also link one of areas of the microgrids to another and connect microgrids with main grids.

Considering the ring topology structure of the microgrid and the power transmission orientations among different areas, the voltage control model of subsystem i with interconnections in the dq-coordinates in terms of Kirchhoff's Curent Law (KCL) and Kirchhoff's Voltage Law (KVL) can be obtained as follows,

$$C_{ti}\dot{V}_{di} = \omega V_{qi} + I_{tdi} - I_{ldi} - I_{di} + \sum_{\substack{j\in N \\ j\neq i}} \xi_{ij} I_{dj} \tag{1}$$

$$C_{ti}\dot{V}_{qi} = -\omega V_{di} + I_{tqi} - I_{lqi} - I_{qi} + \sum_{\substack{j\in N \\ j\neq i}} \xi_{ij} I_{qj} \tag{2}$$

$$L_{ti}\dot{I}_{tdi} = -V_{di} - R_{ti} I_{tdi} + \omega L_{ti} I_{tqi} + U_{di} \tag{3}$$

$$L_{ti}\dot{I}_{tqi} = -V_{qi} - R_{ti} I_{tqi} - \omega L_{ti} I_{tdi} + U_{qi} \tag{4}$$

$$L_i\dot{I}_{di} = V_{di} - \sum_{\substack{j\in N \\ j\neq i}} \xi_{ij} V_{dj} - R_i I_{di} + \omega L_i I_{qi} \tag{5}$$

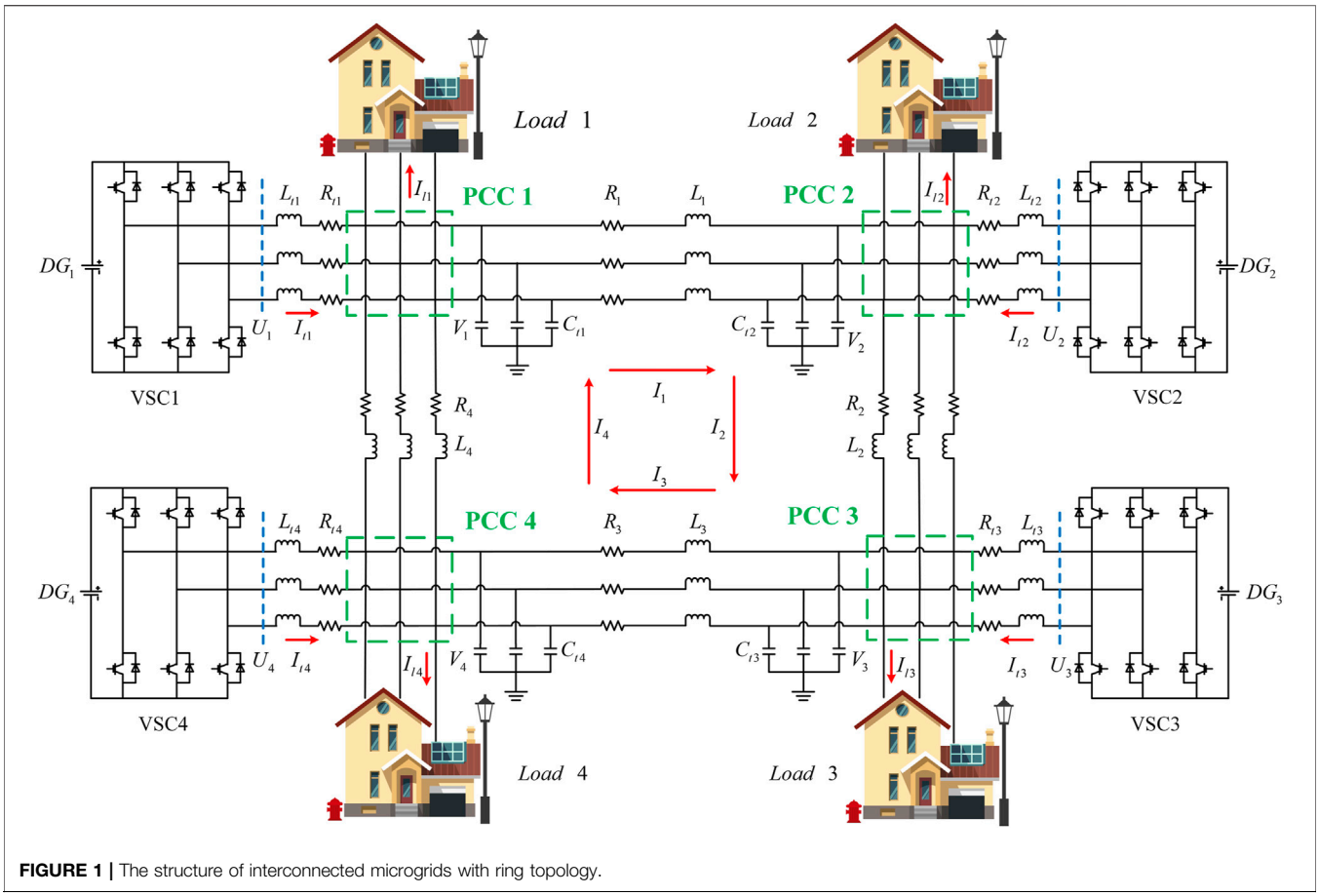

FIGURE 1 | The structure of interconnected microgrids with ring topology.

$$L_i\dot{I}_{qi} = V_{qi} - \sum_{\substack{j\in N \\ j\neq i}} \xi_{ij}V_{qj} - R_iI_{qi} - \omega L_iI_{di} \tag{6}$$

$$y_{1i} = V_{di} \tag{7}$$

$$y_{2i} = V_{qi} \tag{8}$$

where N is the number of DGus in microgrids, L_{ti}, C_{ti} and R_{ti} represent the inductance, capacitance and resistance of the filter connected with the DGu in subsystem i, respectively. The microgrid subsystems in various areas are integrated via interconnecting lines. L_i and R_i are the inductance and resistance of the interconnecting line between subsystem i and the adjacent subsystem. ξ_{ij} is the orientation of the interconnecting line current between subsystem i and subsystem j $(i \neq j)$. $\xi_{ij} = 1$ and $\xi_{ij} = -1$ represent the current flows into and flows out the subsystem i. However, $\xi_{ij} = 0$ represents there is no power exchange between subsystem i and subsystem j. V_{di} and V_{qi} are the direct and quadrature components of the PCC voltage in subsystem i. I_{di} and I_{qi} are the direct and quadrature components of interconnecting line i. I_{tdi}, I_{tqi}, U_{di} and U_{qi} are the direct and quadrature components of the current and voltage generated from DGu in subsystem i. I_{ldi} and I_{lqi} are the direct and quadrature components of local loads. The randomness of the local loads and the power generation intermittence of DGus will cause frequency fluctuation in microgrids. Therefore, we introduce the parameter uncertainties and donate system frequency $\omega = \omega_0 + \Delta\omega$. The matrix form of the dynamic **(1)**–**(8)** can be written as

$$\begin{cases} \dot{x}_i(t) = (A_i + \Delta A_i(t))x_i(t) + B_iu_i(t) + \sum\limits_{\substack{j\in N \\ j\neq i}} E_{ij}x_j(t) + F_id_i(t) \\ y_i(t) = C_ix_i(t) \end{cases} \tag{9}$$

where

$$A_i = \begin{bmatrix} 0 & \omega_0 & \frac{1}{C_{ti}} & 0 & -\frac{1}{C_{ti}} & 0 \\ -\omega_0 & 0 & 0 & \frac{1}{C_{ti}} & 0 & -\frac{1}{C_{ti}} \\ -\frac{1}{L_{ti}} & 0 & -\frac{R_{ti}}{L_{ti}} & \omega_0 & 0 & 0 \\ 0 & -\frac{1}{L_{ti}} & -\omega_0 & -\frac{R_{ti}}{L_{ti}} & 0 & 0 \\ \frac{1}{L_i} & 0 & 0 & 0 & -\frac{R_i}{L_i} & \omega_0 \\ 0 & \frac{1}{L_i} & 0 & 0 & -\omega_0 & -\frac{R_i}{L_i} \end{bmatrix};$$

$$\Delta A_i = \begin{bmatrix} 0 & \Delta\omega & 0 & 0 & 0 & 0 \\ -\Delta\omega & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \Delta\omega & 0 & 0 \\ 0 & 0 & -\Delta\omega & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \Delta\omega \\ 0 & 0 & 0 & 0 & -\Delta\omega & 0 \end{bmatrix};$$

$$E_{ij} = \begin{bmatrix} 0 & 0 & 0 & 0 & \frac{\xi_{ij}}{C_{ti}} & 0 \\ 0 & 0 & 0 & 0 & 0 & \frac{\xi_{ij}}{C_{ti}} \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ -\frac{\xi_{ij}}{L_i} & 0 & 0 & 0 & 0 & 0 \\ 0 & -\frac{\xi_{ij}}{L_i} & 0 & 0 & 0 & 0 \end{bmatrix};$$

$$B_i = \begin{bmatrix} 0 & 0 & \frac{1}{L_{ti}} & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{L_{ti}} & 0 & 0 \end{bmatrix}^T;$$

$$F_i = \begin{bmatrix} -\frac{1}{C_{ti}} & 0 & 0 & 0 & 0 & 0 \\ 0 & -\frac{1}{C_{ti}} & 0 & 0 & 0 & 0 \end{bmatrix}^T;$$

where $x_i = [V_{di} V_{qi} I_{tdi} I_{tqi} I_{di} I_{qi}]^T \in \mathbb{R}^6$ is the state variable vector, $u_i = [U_{di} U_{qi}]^T \in \mathbb{R}^2$ is the control input vector, $d_i(t) = [I_{ldi} I_{lqi}]^T \in \mathbb{R}^2$ is the external disturbance vector and $y_i = [V_{di} V_{qi}]^T \in \mathbb{R}^2$ is the output vector. $A_i \in \mathbb{R}^{6\times6}$, $B_i \in \mathbb{R}^{6\times2}$ and $F_i \in \mathbb{R}^{6\times2}$ are the system matrix, control input matrix and the external disturbance matrix of the ith voltage control model of the microgrid. $\Delta A_i \in \mathbb{R}^{6\times6}$ is a time-varying matrix representing the frequency fluctuation and $E_{ij} \in \mathbb{R}^{6\times6}$ is the interconnection gain matrix consisting of ξ_{ij} and the parameters of interconnecting line i.

Assume that the nominal vector of the state vector in subsystem i is $x_i^* = [V_{di}^* V_{qi}^* I_{tdi}^* I_{tqi}^* I_{di}^* I_{qi}^*]^T$ and $\dot{x}_i^* = 0$. We define the error vector

$$e_i(t) = \begin{bmatrix} V_{di} - V_{di}^* \\ V_{qi} - V_{qi}^* \\ I_{tdi} - I_{tdi}^* \\ I_{tqi} - I_{tqi}^* \\ I_{di} - I_{di}^* \\ I_{qi} - I_{qi}^* \end{bmatrix} = x_i(t) - x_i^*. \tag{10}$$

In view of **(9)**, the corresponding error dynamic model in subsystem i can be expressed by

$$\dot{e}_i(t) = (A_i + \Delta A_i(t))e_i(t) + B_i u_i(t) + \sum_{\substack{j\in N \\ j\neq i}} E_{ij} e_j(t) + f_i(t) \tag{11}$$

where $f_i(t) \triangleq F_i d_i(t) + (A_i + \Delta A_i(t))x_i^* + \sum_{\substack{j\in N \\ j\neq i}} E_{ij} x_j^*$ represents the integration term of exogenous disturbance and parameter uncertainty based on nominal vector, which is Euclidean norm bounded by $\|f_i(t)\| \le \rho_i$, in which

$$\begin{aligned} \|f_i(t)\| &= \left\| F_i d_i(t) + (A_i + \Delta A_i(t))x_i^* + \sum_{\substack{j\in N \\ j\neq i}} E_{ij} x_j^* \right\| \\ &\le \|F_i\|\|d_i(t)\| + \|A_i x_i^*\| + \|a_i x_i^*\| + \left\| \sum_{\substack{j\in N \\ j\neq i}} E_{ij} x_j^* \right\| = \rho_i. \end{aligned} \tag{12}$$

For the later proof proceedings, we introduce the following lemmas, which can be needed to ensure the asymptotic stability of the system.

Lemma 1: (Mnasri and Gasmi, 2011) Consider the following unforced system:

$$\begin{cases} \dot{x} = Ax + H\omega \\ y = Cx \end{cases}$$

This system is regarded as quadratically stable and satisfies the H_∞ norm $\|T_{y\omega}\|_\infty < \gamma$. If there exists a quadratic Lyapunov function $V(x) = x^T P x$, with $P > 0$, then, for all $t > 0$,

$$\dot{V} + y^T y - \gamma^2 \omega^T \omega < 0$$

Lemma 2: (Mnasri and Gasmi, 2011) Let x and y be any vectors with appropriate dimensions. Then, for any scalar $\epsilon > 0$, the following inequality holds:

$$2x^T y \le \epsilon x^T x + \epsilon^{-1} y^T$$

Lemma 3: (Mnasri and Gasmi, 2011) Consider a partitioned symmetric matrix

$$\begin{bmatrix} A & B^T \\ B & C \end{bmatrix},$$

where A and C are square matrices with appropriate dimensions. Then, this matrix is negative define if and only if the matrix A and $C - BA^{-1}B^T$ are negative define.

PROPOSED DECENTRALIZED ADAPTIVE SLIDING MODE VOLTAGE CONTROLLER

The adaptive algorithm can optimize the parameters in controller and the decentralized strategy can improve the control performance. Design proceedings of sliding surface need to consider the stabilizing, tracking and restraining performance of the system. The sliding mode control law usually contains two parts, the switching control law and the equivalent control law. The former one can force the system state to approximate to the sliding surface when it deviates from the surface and the latter one can ensure the system state to keep on the sliding surface when it reaches on the surface.

In order to design decentralized adaptive sliding mode voltage control laws for error dynamic model **(11)**, the following assumptions are introduced.

Assumption 1: All the parameter uncertainty matrices caused by frequency fluctuations are viewed as bounded. That means

$$\|\Delta A_i(t)\| \le a_i. \tag{13}$$

Assumption 2: For each subsystem i,

$$\sum_{j=1, j\neq i}^{N} E_{ji}^T E_{ji} > 0. \tag{14}$$

For improving the dynamic response performance, we define the following integral sliding mode surface as

$$s_i(t) = H_i e_i(t) - \int_0^t H_i[(A_i - B_i K_i)e_i(\tau)]d\tau. \tag{15}$$

where $H_i \in \mathbb{R}^{2\times 6}$ is a constant matrix satisfying $H_i B_i$ is non-singular and $H_i B_i$ is positive for all $i \in N$. $K_i \in \mathbb{R}^{2\times 6}$ is the feedback matrix to be obtained via solving LMIs.

Substituting **Equation 11** into the derivative of sliding surface **(15)** yields

$$\dot{s}_i(t) = H_i\left[(\Delta A_i(t) + B_i K_i)e_i(t) + B_i u_i(t) + \sum_{\substack{j\in N \\ j\neq i}} E_{ij} e_j(t) + f_i(t)\right]. \tag{16}$$

When the state trajectory of the tracking error system arrive and keep on the sliding mode surface, it would satisfy the following equation.

$$s_i(t) = 0 \text{ and } \dot{s}_i(t) = 0 \tag{17}$$

Based on **Equation 17**, equivalent control law can be represented as

$$u_{ieq}(t) = -K_i e_i(t) - (H_i B_i)^{-1} H_i\left[\Delta A_i(t) e_i(t) + \sum_{\substack{j\in N \\ j\neq i}} E_{ij} e_j(t) + f_i(t)\right]. \tag{18}$$

Substituting **Equation (18)** into **(11)**, the sliding mode dynamic equation can further be expressed as

$$\dot{e}_i(t) = \left(A_i - B_i K_i + \widetilde{B}_i \Delta A_i(t)\right)e_i(t) + \widetilde{B}_i \sum_{\substack{j\in N \\ j\neq i}} E_{ij} e_j(t) + \widetilde{B}_i f_i(t) \tag{19}$$

where $\widetilde{B}_i = I - B_i(H_i B_i)^{-1} H_i$. **Equation 19** shows a more complicated tracking error system with parameter uncertainties and external disturbances. In the following procedures, we utilize the Lyapunov theory to analysis the system stability and the tracking performance to the nominal values of currents and voltages in each subsystem. Furthermore, we consider the H_∞ disturbance attenuation performance of interconnected system and design LMIs in terms of an H_∞ norm γ_i.

Theorem 1: Assume that the tracking error system **(19)** satisfies Assumption 1 and Assumption 2. If there exists a feasible solution $X_i > 0$, and R_i satisfies the following LMI **(20)**, then we consider uncertain system **(19)** matches the H_∞ condition.

$$\begin{bmatrix} \Omega_i & \widetilde{B}_i\widetilde{E}_i & \widetilde{B}_i & X_i C_i^T & a_i b_i X_i & X_i \\ * & -I & 0 & 0 & 0 & 0 \\ * & * & -\gamma_i^2 I & 0 & 0 & 0 \\ * & * & * & -I & 0 & 0 \\ * & * & * & * & -\varepsilon_1^{-1} I & 0 \\ * & * & * & * & -\omega_0 & -E_i^{-1} \end{bmatrix} < 0 \tag{20}$$

where $\Omega_i = X_i A_i^T + A_i X_i - B_i R_i - R_i^T B_i^T + \varepsilon_i I$ and $\varepsilon_i > 0$ is a positive scalar. We consider the sliding mode surface

$$s_i(t) = H_i e_i(t) - \int_0^t H_i\left[\left(A_i - B_i R_i X_i^{-1}\right)e_i(\tau)\right]d\tau \tag{21}$$

is asymptotic stable.

Proof: Select the following Lyapunov function for tracking error system **(19)**.

$$V_{1i}(t) = e_i^T(t) P_i e_i(t) \tag{22}$$

Based on an H_∞ performance bound for the closed-loop system **(19)**, one can obtain the following derivative.

$$\begin{aligned} J_i &= \dot{V}_{1i}(t) + y_i^T y_i - \gamma_i^2 f_i^T(t) f_i(t) \\ &= 2\dot{e}_i^T(t) P_i e_i(t) + + y_i^T y_i - \gamma_i^2 f_i^T(t) f_i(t) \\ &= e_i^T(t)\left(A_i^T P_i + P_i A_i - K_i^T B_i^T P_i - P_i B_i K_i\right)e_i(t) \\ &+ 2e_i^T(t)\Delta A_i^T(t)\widetilde{B}_i^T P_i e_i(t) + 2\sum_{\substack{j\in N \\ j\neq i}} e_j^T(t) E_{ij}^T \widetilde{B}_i^T P_i e_i(t) \\ &+ 2f_i^T(t)\widetilde{B}_i^T P_i e_i(t) + e_i^T(t) C_i^T C_i e_i(t) - \gamma_i^2 f_i^T(t) f_i(t) \end{aligned} \tag{23}$$

If $\|\widetilde{B}_i\| \le b_i$ and Assumption one is satisfied, using Lemma 1, we get

$$\begin{aligned} &2e_i^T(t)\Delta A_i^T(t)\widetilde{B}_i^T P_i e_i(t) \\ &\le \varepsilon_i e_i^T(t) P_i^2 e_i(t) + \varepsilon_1^{-1} e_i^T(t)\Delta A_i^T(t)\widetilde{B}_i^T \widetilde{B}_i \Delta A_i(t) e_i(t) \\ &\le e_i^T(t)\left[\varepsilon_i P_i^2 + \varepsilon_1^{-1} a_i^2 b_i^2 I\right]e_i(t). \end{aligned} \tag{24}$$

Consider $\zeta_i^T = [e_i^T, e_{i1}^T, \cdots, e_{iN}^T, f_i^T]$ without e_{ii}^T. **Equation 23** can be rewritten as

$$\begin{aligned} J_i &= \dot{V}_{1i}(t) + y_i^T y_i - \gamma_i^2 f_i^T(t) f_i(t) \\ &= \zeta_i^T \begin{bmatrix} \Lambda_i & P_i\widetilde{B}_i E_{i1} & \cdots & P_i\widetilde{B}_i E_{iN} & P_i\widetilde{B}_i \\ * & -I & 0 & 0 & 0 \\ * & * & \ddots & 0 & 0 \\ * & * & * & -I & 0 \\ * & * & * & * & -\gamma_i^2 I \end{bmatrix}\zeta_i + \sum_{\substack{j\in N \\ j\neq i}} e_j^T e_j - e_i^T E_i e_i, \end{aligned} \tag{25}$$

where $\Lambda_i = A_i^T P_i + P_i A_i - K_i^T B_i^T P_i - P_i B_i K_i + \varepsilon_i P_i^2 + \varepsilon_1^{-1} a_i^2 b_i^2 I + C_i^T C_i + E_i$, and $E_i = \sum E_{ij}^T E_{ij}$ .The system under the equivalent control law is stable, if there is a feasible solution for the following LMI.

$$\begin{bmatrix} \Lambda_i & P_i\widetilde{B}_i E_{i1} & \cdots & P_i\widetilde{B}_i E_{iN} & P_i\widetilde{B}_i \\ * & -I & 0 & 0 & 0 \\ * & * & \ddots & 0 & 0 \\ * & * & * & -I & 0 \\ * & * & * & * & -\gamma_i^2 I \end{bmatrix} < 0 \tag{26}$$

Define $\widetilde{E}_i = (E_{i1}, \cdots, E_{iN})$, $X_i = P_i^{-1}$, and $R_i = K_i X_i$. After pre-multiplying and post-multiplying (26) by $\mathrm{diag}[X_i, I, \cdots, I]$, **Equation 26** can be rewritten as

$$\begin{bmatrix} \widetilde{\Omega}_i & \widetilde{B}_i\widetilde{E}_i & \widetilde{B}_i \\ * & -I & 0 \\ * & * & -\gamma_i^2 I \end{bmatrix} < 0, \tag{27}$$

where $\widetilde{\Omega}_i = X_i A_i^T + A_i X_i - B_i R_i - R_i^T B_i^T + \varepsilon_i I + \varepsilon_i^{-1} a_i^2 b_i^2 X_i^2 + X_i C_i^T C_i X_i + X_i E_i X_i$. Then, the LMI **(20)** is obtained with the method of Lemma three and the proof is completed.In the following section, we design the switching control in terms of adaptive algorithm to restrain the state chattering.

Theorem 2: Design the following controller **(28)** for closed-loop system **(11)** in terms of the feasible solution obtained via **(20)** and the system dynamic is asymptotic stable based on

$$u_i(t) = -K_i e_i(t) - \widetilde{H}_i \widehat{a}_i(t) e_i(t) - \widetilde{H}_i \sum_{\substack{j\in N \\ j\neq i}} E_{ij} e_j(t) - \widehat{\rho}_i(t)\widetilde{H}_i \frac{H_i^T s_i(t)}{\left\| s_i^T(t) H_i \right\|} \tag{28}$$

where $\widetilde{H}_i = (H_i B_i)^{-1} H_i$ and the adaptive laws are

$$\dot{\widehat{a}}_i(t) = q_{i1} s_i^T(t) H_i e_i(t) \tag{29}$$

$$\dot{\widehat{\rho}}_i(t) = q_{i2} \left\| s_i^T(t) H_i \right\| \tag{30}$$

where q_{i1} and q_{i2} are positive parameters.

Proof: Let us consider the following Lyapunov function:

$$V_{2i}(t) = \frac{1}{2} s_i^T(t) s_i(t) + \frac{1}{2q_{i1}} \widetilde{a}_i^2(t) + \frac{1}{2q_{i2}} \widetilde{\rho}_i^2(t) \tag{31}$$

where $\widetilde{a}_i(t) = a_i - \widehat{a}_i(t)$, and $\widetilde{\rho}_i(t) = \rho_i - \widehat{\rho}_i(t)$. Based on **(28)**, **(29)** and **(30)**, its derivative is given by

$$\begin{aligned} \dot{V}_{2i}(t) &= s_i^T(t)\dot{s}_i(t) + \frac{1}{q_{i1}}\widetilde{a}_i(t)\left(-\dot{\widehat{a}}_i(t)\right) + \frac{1}{q_{i2}}\widetilde{\rho}_i(t)\left(-\dot{\widehat{\rho}}_i(t)\right) \\ &= s_i^T(t)\left(H_i\dot{e}_i(t) - H_i(A_i - B_iK_i)e_i(t)\right) \\ &-\widetilde{a}_i(t) s_i^T(t) H_i e_i(t) - \widetilde{\rho}_i(t)\left\| s_i^T(t) H_i \right\| \\ &= s_i^T(t) H_i \Delta A_i(t) e_i(t) - s_i^T(t) H_i \widehat{a}_i(t) e_i(t) \\ &+ s_i^T(t) H_i f_i(t) - \rho_i(t) s_i^T(t) H_i \frac{H_i^T s_i(t)}{\left\| s_i^T(t) H_i \right\|} \\ &-\widetilde{a}_i(t) s_i^T(t) H_i e_i(t) - \widetilde{\rho}_i(t)\left\| s_i^T(t) H_i \right\| \\ &= s_i^T(t) H_i \Delta A_i(t) e_i(t) - a_i s_i^T(t) H_i e_i(t) \\ &+ s_i^T(t) H_i f_i(t) - \rho_i \left\| s_i^T(t) H_i \right\| \\ &\le \left\| \Delta A_i(t) \right\| \left\| s_i^T(t) H_i e_i(t) \right\| - a_i s_i^T(t) H_i e_i(t) \\ &+ \left\| f_i(t) \right\| \left\| s_i^T(t) H_i \right\| - \rho_i \left\| s_i^T(t) H_i \right\| \le 0 \end{aligned} \tag{32}$$

Obviously, the derivative of the Lyapunov function $V_{2i}(t) \le 0$ is verified. That means the system states will reach the designed sliding mode surface in finite time for arbitrary $s_i(t) \neq 0$. Then, the proof is completed.

SIMULATION RESULTS

In this section, the proposed decentralized ASMC strategy is applied in the voltage control of microgrid with ring topology. The microgrid on study contains four DGus ($N = 4$) and the nominal frequency is 60Hz, that means $\omega_0 = 120\pi rad/s$. The electrical parameters of the subsystems and interconnecting lines are concluded in **Table 1** and **Table 2**, respectively.

The intermittence of the renewable generation and the uncertaintiy of the local load in microgrid will influence the power sharing among subsystems and cause frequency fluctuation indirectly. In this case, we consider the frequency fluctuation $\Delta\omega = \sin(1000\pi t)$, the current generated by DGu in subsystem one is increased by 20% at $t = 1$s and recover to the original state at $t = 2$s. Furthermore, the current generated by DGu in subsystem three is reduced by 50% at $t = 1$s and recover to the original state at $t = 2$s. On the contrary, the operation states in subsystem two and four are constant. In order to ensure the power sharing, the voltages of PCCs and currents of interconnecting lines should also change. The reference values of voltages and currents in each subsystem are expressed in **Table 3**.

In the following simulation procedures, we analysis the simulation results and verify the validity of the proposed decentralized ASMC strategies under frequency disturbances and local load uncertainties. Because the system uncertainties are bounded, we can get $a_i \ge 1$, $b_i = 1.7321$. The initial states are $e_1(0) = [5\ -32\ -0.2\ \ 1.5\ \ 1.5]^T$, $e_2(0) = [4.5\ -3.1\ 4\ -0.1\ 1.5\ 2]^T$, $e_3(0) = [4.4\ -3.2\ 1.7\ -0.1\ 2.5\ 1.7]^T$ and $e_4(0) = [4.3\ 3.5\ 3.2\ -0.2\ 1.8\ 2.6]^T$. According to Theroem 1, a couple of feasible solutions can be obtained as follows,

$$H_1 = \begin{bmatrix} 0.0095 & 0 & 0.0095 & 0 & 0.0095 & 0 \\ 0 & 0.0095 & 0 & 0.0095 & 0 & 0.0095 \end{bmatrix},$$

$$H_2 = \begin{bmatrix} 0.0092 & 0 & 0.0092 & 0 & 0.0092 & 0 \\ 0 & 0.0092 & 0 & 0.0092 & 0 & 0.0092 \end{bmatrix},$$

$$H_3 = \begin{bmatrix} 0.0087 & 0 & 0.0087 & 0 & 0.0087 & 0 \\ 0 & 0.0087 & 0 & 0.0087 & 0 & 0.0087 \end{bmatrix},$$

$$H_4 = \begin{bmatrix} 0.0083 & 0 & 0.0083 & 0 & 0.0083 & 0 \\ 0 & 0.0083 & 0 & 0.0083 & 0 & 0.0083 \end{bmatrix},$$

$$K_1 = \begin{bmatrix} 9214.5902 & 0 & 2416.5281 & 0 & -6560.1562 & 0 \\ 0 & 9214.5902 & 0 & 2416.5281 & 0 & -6560.1562 \end{bmatrix},$$

$$K_2 = \begin{bmatrix} 8795.6529 & 0 & 2451.4103 & 0 & -8789.3652 & 0 \\ 0 & 8795.6529 & 0 & 2451.4103 & 0 & -8789.3652 \end{bmatrix},$$

$$K_3 = \begin{bmatrix} 10225.614 & 0 & 2484.2629 & 0 & -12448.709 & 0 \\ 0 & 10225.614 & 0 & 2484.2629 & 0 & -12448.709 \end{bmatrix},$$

$$K_4 = \begin{bmatrix} 8079.6892 & 0 & 2208.2329 & 0 & 5562.70592 & 0 \\ 0 & 8079.6892 & 0 & 2208.2329 & 0 & 5562.70592 \end{bmatrix}.$$

In order to restrain the mismatched uncertainties and external disturbances, for tracking error system **(11)**, the adaptive parameters are selected as

$$q_{11} = 1.2,\ q_{21} = 0.9,\ q_{31} = 1.1,\ q_{41} = 1.5$$

TABLE 1 | Electrical parameters of the microgrid.

Subsystems	Filter parameters		Capacitance
	R_{ti} (mΩ)	L_{ti} (mH)	C_{ti} (μF)
Subsystem 1	40.2	9.5	62.86
Subsystem 2	38.7	9.2	62.86
Subsystem 3	34.6	8.7	62.86
Subsystem 4	31.8	8.3	62.86

TABLE 2 | Electrical parameters of the interconnecting lines.

Line impedance Z_i	R_i (Ω)	L_i (mH)
Line 1	7.6	111.9
Line 2	8	140
Line 3	7.3	165
Line 4	7.8	190

TABLE 3 | The reference values of state variables in each subsystem.

Time(s)	Subsystem1				Subsystem2			
	Voltages (V)		Currents (A)		Voltages (V)		Currents (A)	
	V^*_{d1}	V^*_{q1}	I^*_{td1}	I^*_{tq1}	V^*_{d2}	V^*_{q2}	I^*_{td2}	I^*_{tq2}
0–1	390	−0.5	50.2	−10.6	370	−1	100.1	−6.8
1–2	385	−0.5	70.2	−10.6	370	−1	100.1	−6.8
2–3	390	−0.5	50.2	−10.6	370	−1	100.1	−6.8
Time(s)	**Subsystem3**				**Subsystem4**			
	Voltages (V)		Currents (A)		Voltages (V)		Currents (A)	
	V^*_{d3}	V^*_{q3}	I^*_{td3}	I^*_{tq3}	V^*_{d4}	V^*_{q4}	I^*_{td4}	I^*_{tq4}
0–1	360	−0.6	40.1	−1.8	350	−1	80.1	−10.4
1–2	365	−0.6	20.1	−1.8	350	−1	80.1	−10.4
2–3	360	−0.6	40.1	−1.8	350	−1	80.1	−10.4

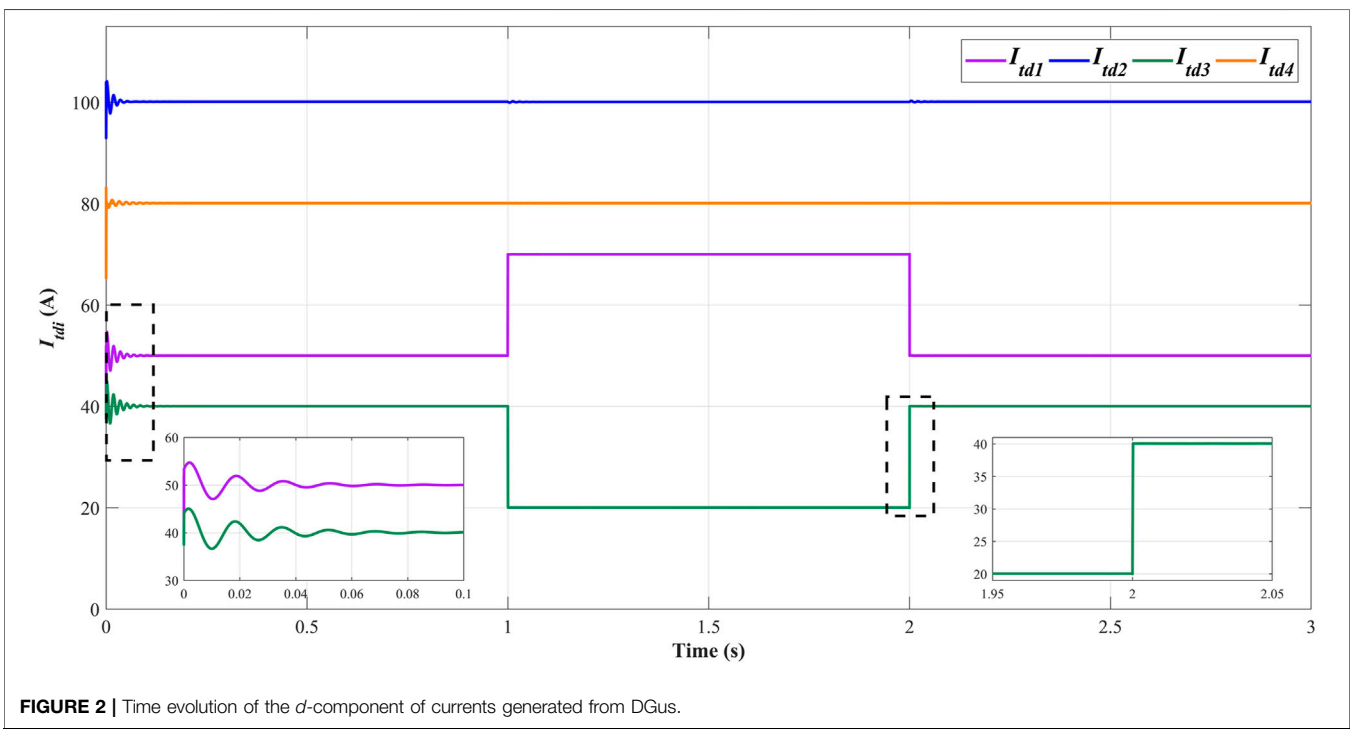

FIGURE 2 | Time evolution of the *d*-component of currents generated from DGus.

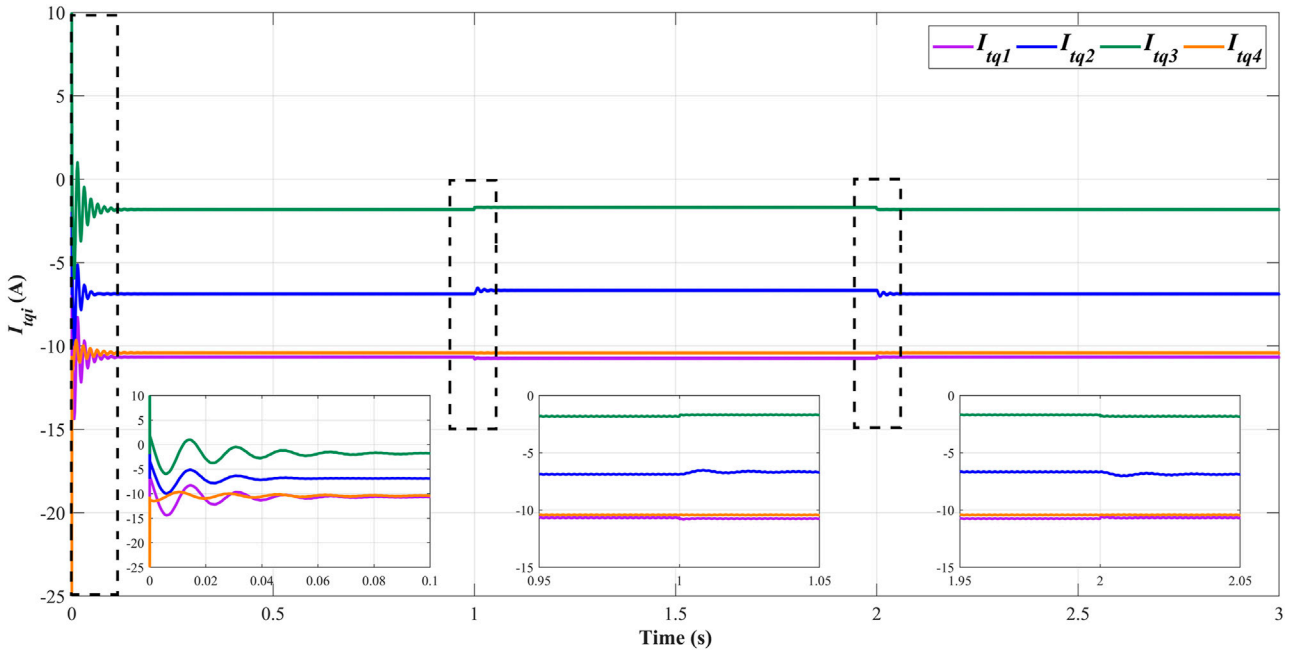

FIGURE 3 | Time evolution of the *q*-component of currents generated from DGus.

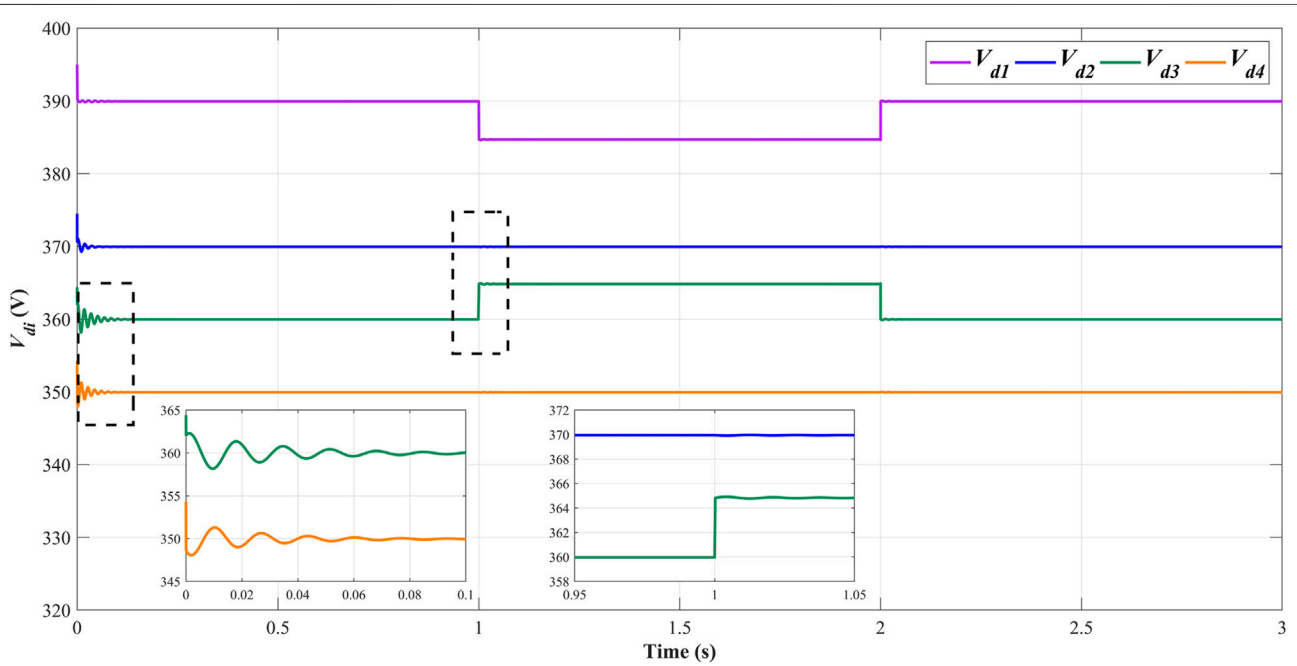

FIGURE 4 | Time evolution of the *d*-component of voltages of PCCs.

$$q_{12} = 1.0,\ q_{22} = 3.2,\ q_{32} = 3.7,\ q_{42} = 5.74$$

The voltages and currents of the islanded microgrid demonstrate superior robustness and tracking error performance under the proposed decentralized ASMC strategy. The time evolutions of the *dq*-components of the currents generated from each DGu are depicted in **Figure 2** and **Figure 3** respectively in the case of load currents mutation. I_{ld1} increases 20 A at $t = 1$s and I_{td1} increases 20 A synchronously, while I_{tq1} does not change significantly. Similarly, the current generated from DGu3 varies with the local load current in subsystem 3, and the current dynamic

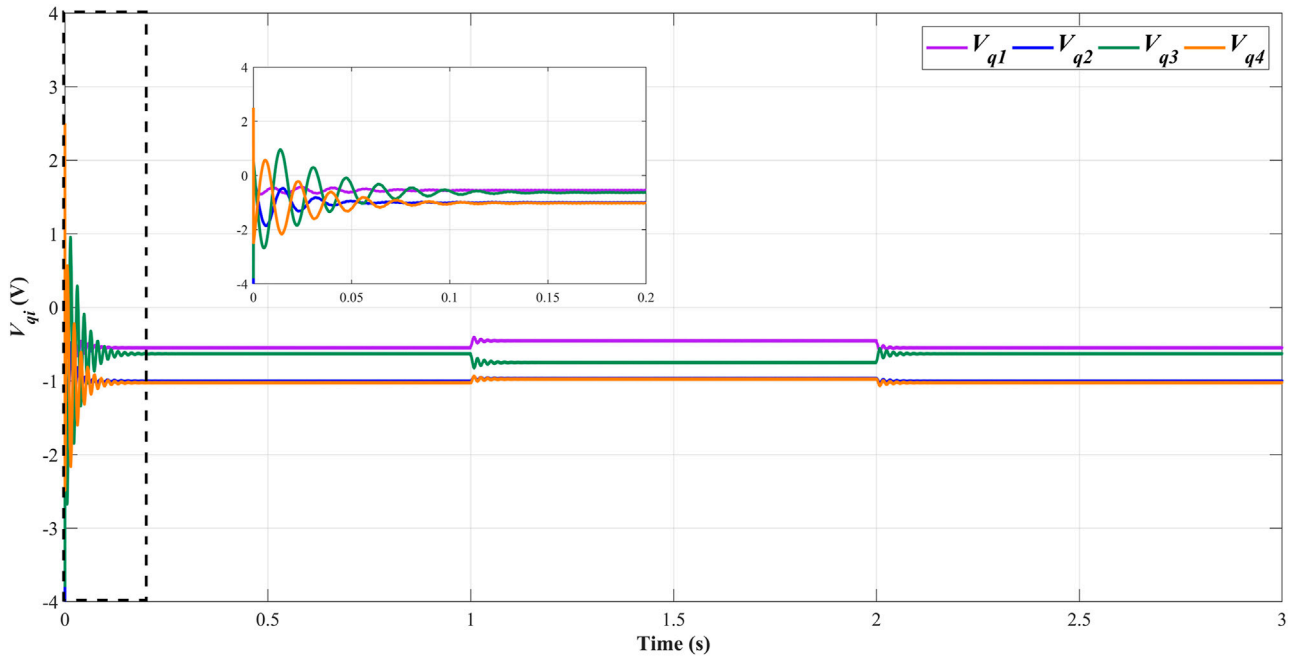

FIGURE 5 | Time evolution of the *q*-component of voltages of PCCs.

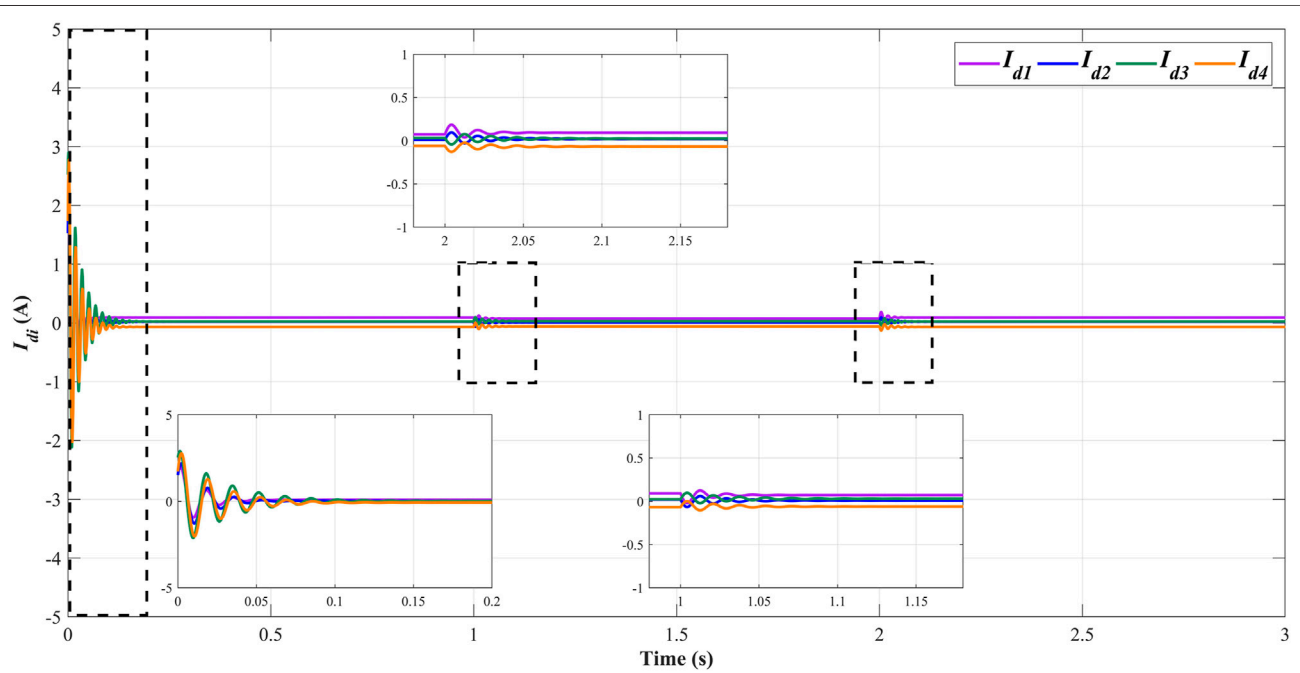

FIGURE 6 | Time evolution of the *d*-component of currents of interconnecting lines.

curves illustrate excellent robustness when local load current mutates.

The time evolutions of the *dq*-components of the PCC voltages of each subsystem are shown in **Figure 4** and **Figure 5**. Considering the load current variations in the interconnected microgrid system, V_{d1} is decreased by 5 V and V_{d3} is increased by 5 V. Meanwhile, I_{d1} provided by subsystem one is reduced and I_{d3} provided to subsystem four is improved. Then, the power sharing among each subsystem can be ensured. As shown in **Figure 4**, there exist deviations between initial voltage values and nominal values, but the system voltages can also track to the nominal values and maintain stable. When the *d*-components of the

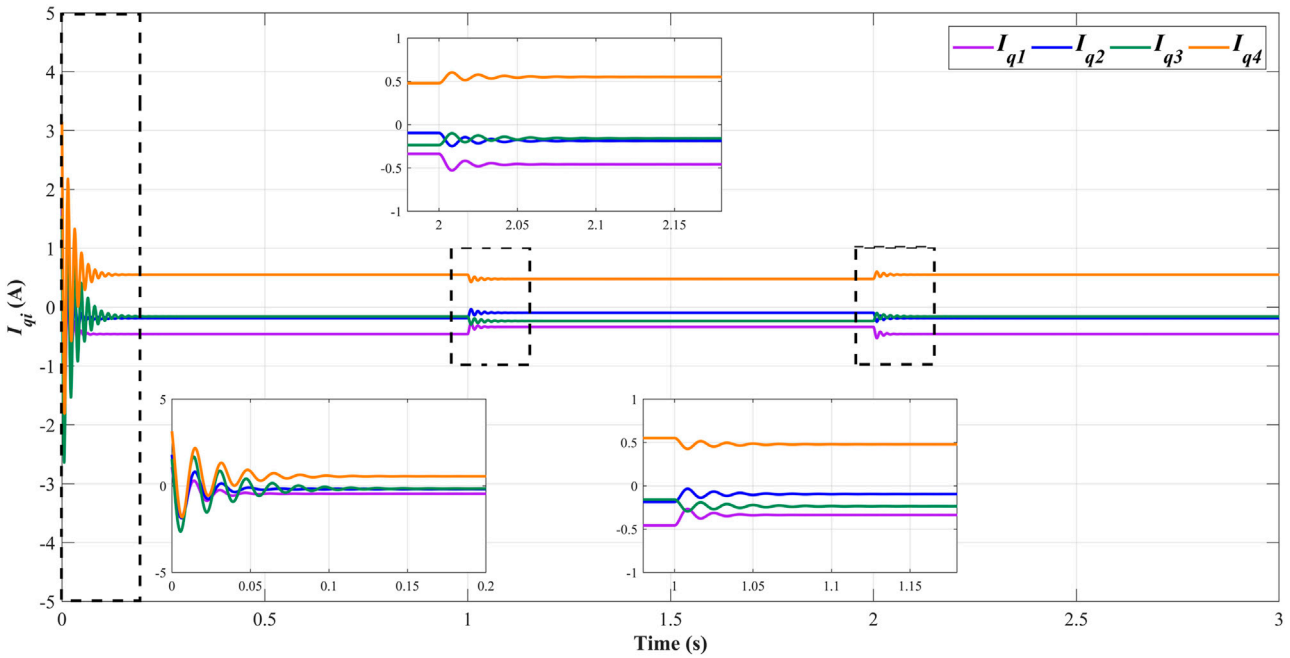

FIGURE 7 | Time evolution of the *q*-component of currents of interconnecting lines.

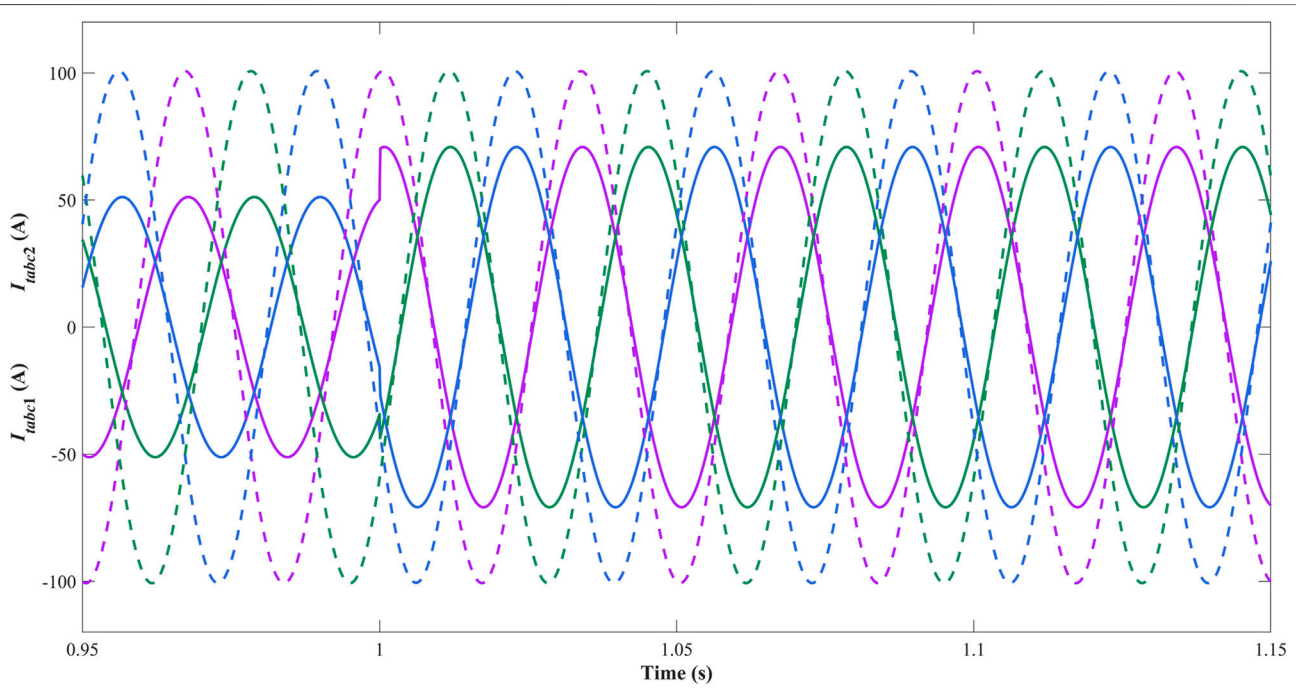

FIGURE 8 | Three-phase waveforms of currents generated from DGu1 and DGu2.

voltages change in small scales, the *q*-components of the voltages fluctuate and remain stable in a short time.

The time evolutions of the *dq*-components of the currents of interconnecting lines are represented in **Figure 6** and **Figure 7**. Under the influence of mutations of load currents and PCC voltages, the currents of interconnecting lines are only influenced by the voltages of PCCs and do not deviate from the reference values appreciably, which illustrates the remarkable robustness of proposed ASMC strategy against the external disturbances and mismatched uncertainties. That also means the changes of local

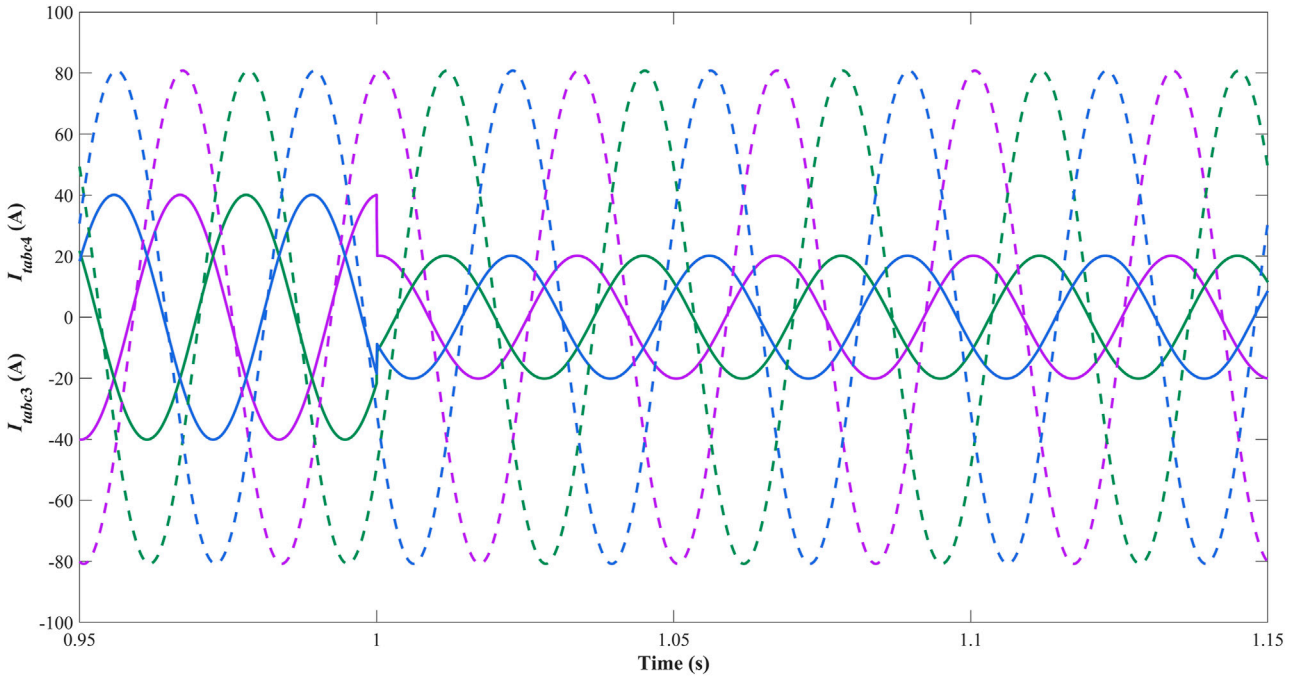

FIGURE 9 | Three-phase waveforms of currents generated from DGu3 and DGu4.

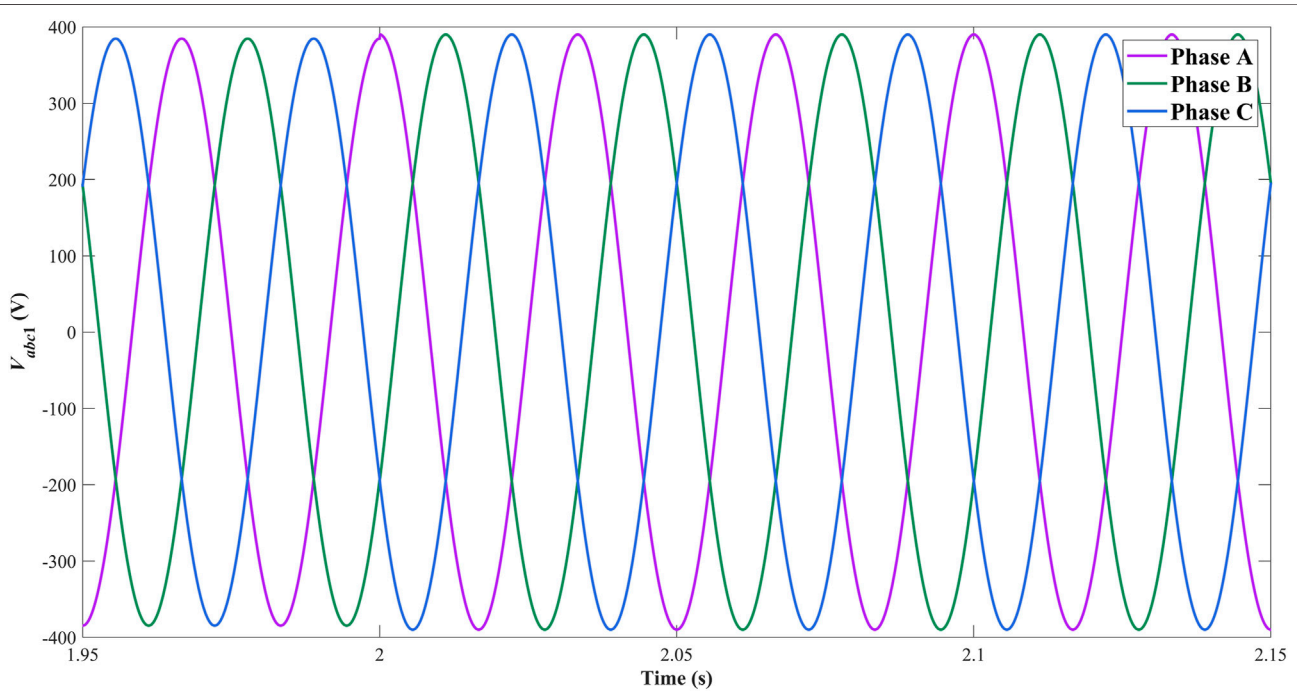

FIGURE 10 | Three-phase waveforms of PCC1 voltage.

load currents will not influence the stability of the general microgrid. Notably, the current orientation of line four is opposite compared with the currents of other lines because the voltage of PCC four is lower than that of PCC1.

Figure 8 and **Figure 9** depict the three-phase waveforms of currents generated from DGus. **Figure 10** and **Figure 11** depict the three-phase waveforms of voltages of PCCs in subsystem one and subsystem 3. In islanded mode, the decentralized controllers are

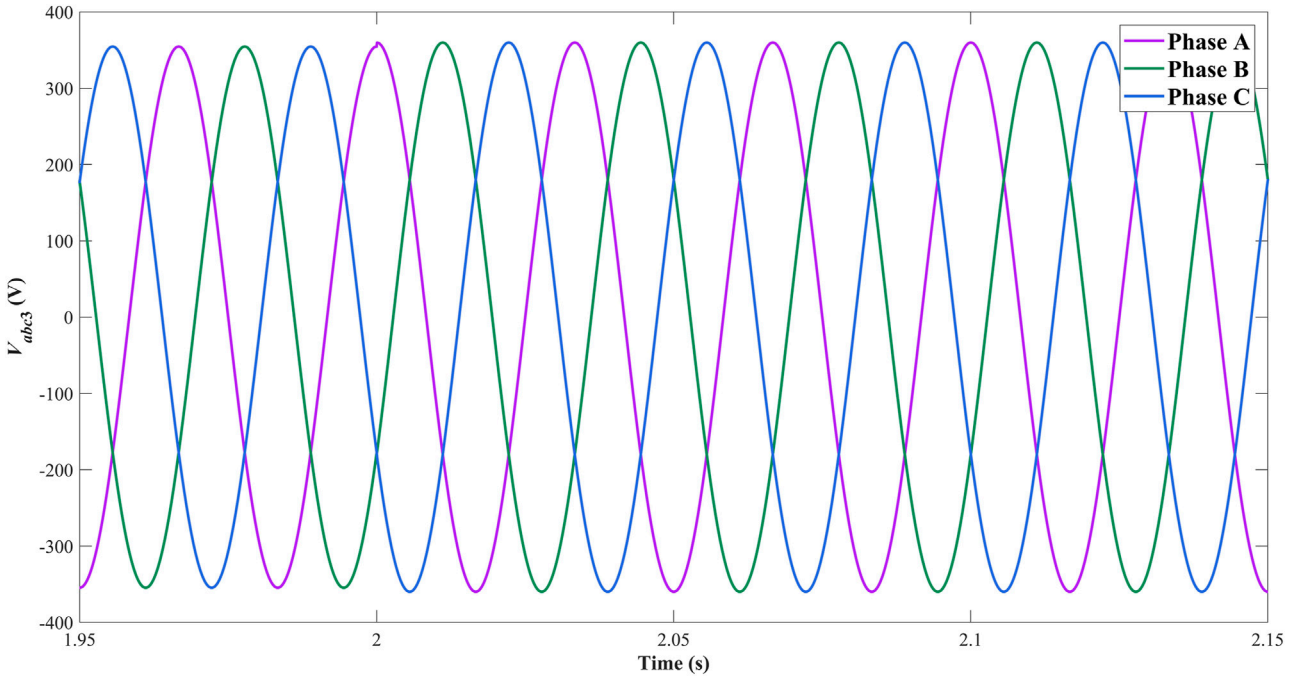

FIGURE 11 | Three-phase waveforms of PCC3 voltage.

mainly designed to standardize the voltage and current waves and improve the power quality of AC microgrid. In **Figure 8** and **Figure 9**, the amplitude of three-phase wave of the current generated from DGu1 increases 20 A and that of DGu3 decreases 20 A at $t = 1$s, which matches with the variations of load currents. In **Figure 10** and **Figure 11**, it is obvious that the three-phase waves of the voltages of PCC1 and PCC3 maintain in standard waveform and the effectiveness and reliability can be proved adequately.

CONCLUSION

In this paper, the chattering restraint issues for voltage control of the islanded microgrid with a ring topology structure have been solved via decentralized adaptive sliding mode control strategies. The constructed tracking error system with interconnections has depicted the interaction among each subsystem appropriately. The control matrices in sliding mode surfaces have been obtained via solving the LMIs, which combine the H_∞ attenuation performance of the system external disturbances with the asymptotic stabilities of integral sliding mode surfaces. The controller parameters have been optimized by means of adaptive algorithms. The simulation results have illustrated the effectiveness of the proposed decentralized ASMC strategies. In future, further research will be extended to the nonlinear and time-delay system.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

AUTHOR CONTRIBUTIONS

HJ was responsible for the main idea and writing work of this paper. YZ and JH were responsible for the mathematical derivation. MW was responsible for the simulation part.

FUNDING

Doctoral Scientific Research Foundation of Liaoning Province (2020-BS-181). Liaoning Revitalization Talents Program (XLYC1907138). Natural Science Foundation of Liaoning Province (2019-MS-239). Key R&D Program of Liaoning Province (2020JH2/10300101).

REFERENCES

Avelar, H. J., Parreira, W. A., Vieira, J. B., de Freitas, L. C. G., and Coelho, E. A. A. (2012). A State Equation Model of a Single-phase Grid-Connected Inverter Using a Droop Control Scheme with Extra Phase Shift Control Action. *IEEE Trans. Ind. Electron.* 59, 1527–1537. doi:10.1109/TIE.2011.2163372

Beerten, J., and Belmans, R. (2013). Analysis of Power Sharing and Voltage Deviations in Droop-Controlled DC Grids. *IEEE Trans. Power Syst.* 28, 4588–4597. doi:10.1109/TPWRS.2013.2272494

Chen, Z., Luo, A., Wang, H., Chen, Y., Li, M., and Huang, Y. (2015). Adaptive Sliding-Mode Voltage Control for Inverter Operating in Islanded Mode in Microgrid. *Int. J. Electr. Power Energ. Syst.* 66, 133–143. doi:10.1016/j.ijepes.2014.10.054

Cucuzzella, M., Incremona, G. P., and Ferrara, A. (2017). Decentralized Sliding Mode Control of Islanded AC Microgrids with Arbitrary Topology. *IEEE Trans. Ind. Electron.* 64, 6706–6713. doi:10.1109/TIE.2017.2694346

Divshali, P. H., Alimardani, A., Hosseinian, S. H., and Abedi, M. (2012). Decentralized Cooperative Control Strategy of Microsources for Stabilizing Autonomous VSC-Based Microgrids. *IEEE Trans. Power Syst.* 27, 1949–1959. doi:10.1109/TPWRS.2012.2188914

Eren, S., Pahlevani, M., Bakhshai, A., and Jain, P. (2015). An Adaptive Droop DC-Bus Voltage Controller for a Grid-Connected Voltage Source Inverter with LCL Filter. *IEEE Trans. Power Electron.* 30, 547–560. doi:10.1109/TPEL.2014.2308251

Esparza, M., Segundo, J., Nunez, C., Wang, X., and Blaabjerg, F. (2017). A Comprehensive Design Approach of Power Electronic-Based Distributed Generation Units Focused on Power-Quality Improvement. *IEEE Trans. Power Deliv.* 32, 942–950. doi:10.1109/TPWRD.2016.2584616

Hu, J., Nian, H., Hu, B., He, Y., and Zhu, Z. Q. (2010). Direct Active and Reactive Power Regulation of DFIG Using Sliding-Mode Control Approach. *IEEE Trans. Energ. Convers.* 25, 1028–1039. doi:10.1109/TEC.2010.2048754

Kabalan, M., Singh, P., and Niebur, D. (2017). Large Signal Lyapunov-Based Stability Studies in Microgrids: A Review. *IEEE Trans. Smart Grid* 8, 2287–2295. doi:10.1109/TSG.2016.2521652

Karimi, H., Davison, E. J., and Iravani, R. (2010). Multivariable Servomechanism Controller for Autonomous Operation of a Distributed Generation Unit: Design and Performance Evaluation. *IEEE Trans. Power Syst.* 25, 853–865. doi:10.1109/TPWRS.2009.2031441

Liu, J., Gao, Y., Luo, W., and Wu, L. (2017). Takagi-Sugeno Fuzzy-Model-Based Control of Three-Phase AC/DC Voltage Source Converters Using Adaptive Sliding Mode Technique. *IET Control. Theor. Appl.* 11, 1255–1263. doi:10.1049/iet-cta.2016.0689

Mahmoud, M. S., Azher Hussain, S., and Abido, M. A. (2014). Modeling and Control of Microgrid: An Overview. *J. Franklin Inst.* 351, 2822–2859. doi:10.1016/j.jfranklin.2014.01.016

Mi, Y., Song, Y., Fu, Y., and Wang, C. (2020). The Adaptive Sliding Mode Reactive Power Control Strategy for Wind-Diesel Power System Based on Sliding Mode Observer. *IEEE Trans. Sustain. Energ.* 11, 2241–2251. doi:10.1109/TSTE.2019.2952142

Mi, Y., Zhang, H., Fu, Y., Wang, C., Loh, P. C., and Wang, P. (2019). Intelligent Power Sharing of DC Isolated Microgrid Based on Fuzzy Sliding Mode Droop Control. *IEEE Trans. Smart Grid* 10, 2396–2406. doi:10.1109/TSG.2018.2797127

Mnasri, C., and Gasmi, M. (2011). LMI-Based Adaptive Fuzzy Integral Sliding Mode Control of Mismatched Uncertain Systems. *Int. J. Appl. Maths. Comput. Sci.* 21, 605–615. doi:10.2478/v10006-011-0047-5

Rui, W., Qiuye, S., Dazhong, M., and Xuguang, H. (2020). Line Impedance Cooperative Stability Region Identification Method for Grid-Tied Inverters under Weak Grids. *IEEE Trans. Smart Grid* 11, 2856–2866. doi:10.1109/TSG.2020.2970174

Sahoo, S. K., Sinha, A. K., and Kishore, N. K. (2018). Control Techniques in AC, DC, and Hybrid AC-DC Microgrid: A Review. *IEEE J. Emerg. Sel. Top. Power Electron.* 6, 738–759. doi:10.1109/JESTPE.2017.2786588

Sefa, I., Altin, N., Ozdemir, S., and Kaplan, O. (2015). Fuzzy PI Controlled Inverter for Grid Interactive Renewable Energy Systems. *IET Renew. Power Generation* 9, 729–738. doi:10.1049/iet-rpg.2014.0404

Vandoorn, T. L., Ionescu, C. M., De Kooning, J. D. M., De Keyser, R., and Vandevelde, L. (2013). Theoretical Analysis and Experimental Validation of Single-Phase Direct Versus Cascade Voltage Control in Islanded Microgrids. *IEEE Trans. Ind. Electron.* 60, 789–798. doi:10.1109/TIE.2012.2205362

Wang, R., Sun, Q., Hu, W., Li, Y., Ma, D., and Wang, P. (2021). SoC-Based Droop Coefficients Stability Region Analysis of the Battery for Stand-Alone Supply Systems with Constant Power Loads. *IEEE Trans. Power Electron.* 36, 7866–7879. doi:10.1109/TPEL.2021.3049241

Wang, R., Sun, Q., Ma, D., and Liu, Z. (2019). The Small-Signal Stability Analysis of the Droop-Controlled Converter in Electromagnetic Timescale. *IEEE Trans. Sustain. Energ.* 10, 1459–1469. doi:10.1109/TSTE.2019.2894633

Zeb, K., Islam, S. U., Din, W. U., Khan, I., Ishfaq, M., Busarello, T. D. C., et al. (2019). Design of Fuzzy-PI and Fuzzy-Sliding Mode Controllers for Single-Phase Two-Stages Grid-Connected Transformerless Photovoltaic Inverter. *Electronics* 8, 520–539. doi:10.3390/electronics8050520

Zhou, B., Zou, J., Yung Chung, C., Wang, H., Liu, N., Voropai, N., et al. (2021). Multi-Microgrid Energy Management Systems: Architecture, Communication, and Scheduling Strategies. *J. Mod. Power Syst. Clean Energ.* 9, 463–476. doi:10.35833/MPCE.2019.000237

Zou, H., Mao, S., Wang, Y., Zhang, F., Chen, X., and Cheng, L. (2019). A Survey of Energy Management in Interconnected Multi-Microgrids. *IEEE Access* 7, 72158–72169. doi:10.1109/ACCESS.2019.2920008

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Energy Research

ORIGINAL RESEARCH
published: 07 September 2021
doi: 10.3389/fenrg.2021.730632

Power Dispatching of Transportable Energy Storage System for Post-disaster Restoration Scheme of Port: The AES-Based Joint Restoration Scheme

Wenjia Xia[1], *Qihe Shan*[1]*, *Fei Teng*[2] *and Tieshan Li*[1,3]

[1]Maritime Big Data and Artificial Intelligent Application Centre, Navigation College, Dalian Maritime University, Dalian, China, [2]Marine Electrical Engineering College, Dalian Maritime University, Dalian, China, [3]College of Automation Engineering, University of Electronic Science and Technology, Chendu, China

In recent years, the rapid development of artificial intelligence, Big Data, Cloud Computing, etc., rapidly develops, synergy development of the transportation-energy-information based triple play has been accelerated. The all-electric ship (AES) is satisfied with the demand for both the unreasonable resource configuration of port energy system and low energy efficiency caused by sudden disasters. As a power carrier unit, AES has great potential in improving the resistance of port power distribution network for large-area blackout. Therefore, in this paper, a joint post-disaster restoration scheme for AESs and distribution generation scheduling in ports' microgrid (MG) is proposed to minimize the cost of post-disaster restoration including customer interruption cost and AES operation cost. The time-space network (TSN) model which describes the scheduling of AESs is applied to compare the grid resilience and flexibility of AES-based joint dispatching and only relying on distributed generation equipment in post-disaster restoration. The joint scheduling problem is formulated as a mixed-integer linear programming. The proposed model and scheme are tested in a system with 3 ports and 4 AESs verified the resilience of the port power distribution network with AESs.

Keywords: AES, port microgrid, post-disaster restoration, energy scheduling, transportable energy storage system

OPEN ACCESS

Edited by:
Qiuye Sun,
Northeastern University, China

Reviewed by:
Yue Wu,
Southwest Jiaotong University, China
Chaoke Gong,
Dalian University of Technology, China

***Correspondence:**
Qihe Shan
shanqihe@163.com

Specialty section:
This article was submitted to Smart Grids, a section of the journal Frontiers in Energy Research

Received: *25 June 2021*
Accepted: *23 August 2021*
Published: *07 September 2021*

Citation:
Xia W, Shan Q, Teng F and Li T (2021) Power Dispatching of Transportable Energy Storage System for Post-disaster Restoration Scheme of Port: The AES-Based Joint Restoration Scheme. Front. Energy Res. 9:730632. doi: 10.3389/fenrg.2021.730632

1 INTRODUCTION

With the acceleration of the global economic integration and the international economic activity becoming more and more frequent. Port as a national window opening to the outside world plays an increasingly important role. It is essential to promote coastal cities and has been regarded as the growth engine of a region or national economy (Li and Dong, 2014), especially in the regional economy of coastal cities. According to the data provided by the 2010 Shipping Statistical Yearbook, Shanghai port is the largest port in the world. In the same year, China accounted for six of the top ten ports in the world (Wang, 2011). Disasters of port could have far-reaching consequences for our country.

Meanwhile, China is one of the few countries which suffer serious loss from marine disasters (Wang, 2005). And the seaport is one of the hazard bearing bodies, in which large power facilities such as container handling machinery, transporting machine and service loads are damaged in

different degrees, the electricity supply is interrupted, the application of intelligent diagnosis method in power transformers are still fall short (Yang et al., 2021) (Zhang et al., 2021). The post-disaster restoration is hindered which seriously threaten the lives of people's and hard to reduce the economic loss. With increasingly abundant power equipment and high reliance of electricity, power system, as an important infrastructure related to security and economy, not only needs to meet the reliable operation in normal environment, but also maintains the requisites in the extreme condition (Ren et al., 2020). Emergency power supply system is essential.

Emergency power supply system is an independent generation device, which supplies electricity in case of unexpected power outage, increasing the resilience of microgrid, and protects life and property from the impact of power failure. Considerable progress has been made with respect to the utilization of stationary sources in distributed systems (DSs) restoration (Chanda and Srivastava, 2016) (Panteli et al., 2017). Distributed generators (DGs) are widely adopted in the restoration when the DS is isolated from the main grid (Wang and Wang, 2015). More scholars turn to the application of transportable energy storage system for post-disaster restoration.

As the green transportation, clean energy, artificial intelligence and Internet of Things technology (IoT Tech) rapidly develops, synergy development of the transportation-energy-information based triple play has been accelerated, and gives impetus to modernization of infrastructure. The more flexible mobile emergency power bringing new challenges and opportunities for power system recovery has been developed rapidly (Zhou et al., 2017). With the typical transportable energy storage system, e.g., electric vehicle, retention increasing dramatically year by year, V2G technology, self-driving and other relevant techniques having mature, the development tendency of the application of transportable energy storage system in electric power safeguard in the future has realized (Khodayar et al., 2013) (Chen et al., 2019). In 2000, Kempton and others first analyzed the feasibility of electric vehicles as distributed energy storage by coordinating the charging/discharging process of electric vehicles. For transmission networks, electric vehicles are applied to realize load shifting to improve the operation characteristics of grid (Lv, 2018). In recent years, the United States, Denmark, Japan and other developed countries and regions have vigorously promoted the relevant research. Hence, transportable sources, e.g., electric vehicles, have great potential to enhance the resilience through optimal scheduling among multiple islands within the DS.

AESs are regarded as transportable energy storage systems. Under the dual wheel drive of power security and environmental protection, AES with characteristics of clean, environmental protection and high energy efficiency continues to thrive in recent years, which is a kind of new energy ship with wide application prospects. In the port shore power stations, AESs are generally the user's load. The rise of reverse charging technology promotes the transformation of AES from load to distributed power source, realizes the bi-directional interaction between ship grid and port grid. It leave the possibility for AESs to enhance the resilience of port's microgrid.

In comparison to the fixed emergency power supply device, AESs' flexibility is beyond doubt as a transport. AESs serve as a bridge for energy dispatching between ports, realizing flexible energy management as well as meeting the energy demand of post-disasters port's microgrid which assist the post-disaster restoration. As a mobile power storage system, AESs play a strong role in power supply in the emergency of post-disasters.

AESs obviously have more advantages than electric vehicles in the specific environment of port grid. Against the electric vehicles with the identical flexibility, AESs have strong power storage capacity and diversified energy utilization. Namely, AESs' single charging/discharging provide more power accommodation/energy supplement for the post-disaster port microgrid.

Except for the decided advantage in energy storage capacity, the AES is more powerful than electric vehicle in integrated utilization of multiple energy sources. Due to the limitation of cubage and other factors, the energy utilization types of electric vehicles are relatively single. However, the AESs utilize multi-kinds of energy simultaneously, e.g. solar resources, wind power, tide energy and other clean energy. What's more, only can electric vehicles realize the energy dispatching among emergency power sources in the port, and cannot achieve the energy transportation management among ports. The AESs not only carry out dispatching inside the port by using the berthing position, but also dispatch energy among ports. The AESs highly accords with the demand of the post-disaster restoration of the ports' microgrid, which is the excellent choice of the port uninterrupted power supply.

In recent years, the rapid development of artificial intelligence, big data, cloud computing and other related information technology has brought new strategies for energy interconnection and optimal operation which provides opportunities to the unreasonable resource allotment and low resource-efficient of port energy system caused by disasters. In this context, the AES-based joint energy dispatching comes into being. As a kind of carrier, AES building a bridge between the affected port and its neighbor grid realizes the electricity-centered energy interconnection of multiple ports. The AES-based joint scheduling build a foundation for the unified dispatching of multi-port grids and AESs' energy storage system, maximize energy utilization, minimize the cost of post-disaster restoration. Meanwhile, adjustment of using energy from high-carbon to low-carbon is consilience of the current development theory and IMO energy conservation and environmental protection requirements (Fang et al., 2020).

In this paper, the research problems illustrate as follows: the connection of port grid and the main grid is fault after disasters, leading to a self-sufficient islanded microgrid of port. When a port is hit particularly hard, distributed generator are insufficient to meet the energy requirements, and thus AESs can be aggregated as a wide range of uninterruptible power sources for service restoration as far as possible. In this context, a joint post-disaster restoration scheme is proposed, in which AESs are dynamically scheduled in coordination with port restoration through microgrids, to minimize the total cost in the wake of disasters. Specifically, the islanded microgrid and AESs are described by using the temporal-spatial network model, and

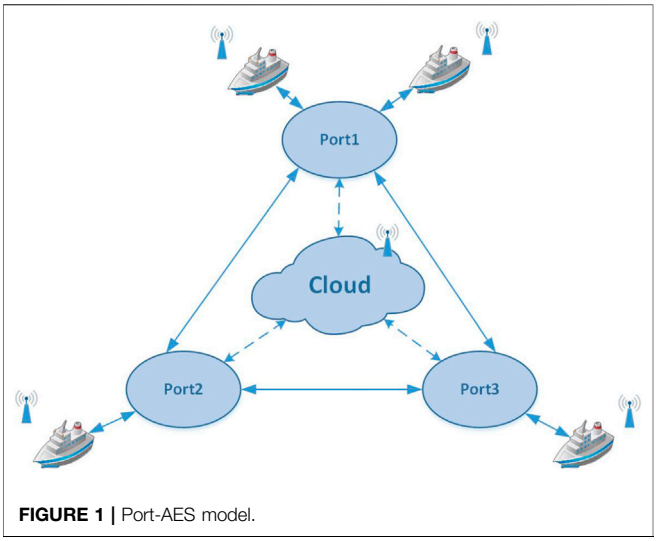

FIGURE 1 | Port-AES model.

the mathematical model of AES operation is established to optimize the power dispatching scheme.

Main contributions are concluded as follows:

1) This paper proposes a joint post-disaster restoration scheme for berthing AESs and distribution generation in multi-port microgrid to minimize the cost of post-disaster restoration.
2) In order to evaluate the necessity and feasibility of AES-based joint scheduling, the mathematical model of restoration cost of the affected ports is established and resolve through MILP mean.

The remainder of the paper consists of four sections. **Section 2** illustrates the system architecture in detail. **Section 3** presents the modeling of AESs, ports and joint post-disaster restoration scheme. **Section 4** provides the numerical results and analysis, and the paper is summarized in **Section 5**.

2 JOINT POST-DISASTER RESTORATION SCHEME

In the aftermath of the extreme events, ports' microgrid has being place on island mode (Parise et al., 2016). AESs, as the power inter-connection bridge of adjacent ports, transfer the electricity to the post-disaster port. The total data, including AESs' position and ports' state, are aggregated to the cloud to work out the dispatching plan to provide guidance for AESs move as shown in **Figure 1**.

AESs join in the post-disaster restoration, increasing the resilience of microgrid of the port. For the sake of the best restoration scheme of port microgrid, the mathematical formulation of the joint post-disaster restoration scheme with the AESs is presented in this section, aiming at both minimizing the cost of joint post-disaster recovery and maximizing the efficiency of microgrid restoration. Furthermore, post-disaster recovery is characterized by the power recovery of local load. The customer interruption cost is adopted by distinguishing the load priority. The objective function and port operation constraints sets are described as follows.

2.1 Objective Function

The objective function considers two aspects: customer interruption cost and AES operation cost. AES part includes three aspects: AES transportation cost, DG generation cost and loss cost of shipborne battery, as follow (Yao et al., 2018):

$$\min \sum_{t\in T}\Bigg[\sum_{m\in M} W_m\left(PD_m^t - PD_{r,m}^t\right)\Delta T + \sum_{m\in M} C_{gen,m}P_{DG,m}^t\Delta T + \sum_{\omega\in\Omega} C_{tran,\omega}\sum_{(m,u)\in Z,m\neq u}\zeta_{\omega,mu}^t + \sum_{\omega\in\Omega} C_{bat,\omega}\sum_{m\in M}\left|P_{ch,\omega m}^t + P_{dch,\omega m}^t\right|\Delta T\Bigg] \tag{1}$$

The first term of the objective **Eq. 1** $\sum_{m\in M} W_m(PD_m^t - PD_{r,m}^t)\Delta T$ represents the sum of customer interruption costs at time t. W_m is unit interruption cost for load at port m, PD_m^t is load in port m in time span t and $PD_{r,m}^t$ is load restored at port m in time span t. The second term $\sum_{m\in M} C_{gen,m}P_{DG,m}^t\Delta T$ represents the power generation cost of the shipborne generator of the AES ω in time span t. $C_{gen,m}$ is unit generation cost for the DGs in port m. The last portions are associated with AES, $C_{tran,\omega}$ and $C_{bat,\omega}$ are unit transportation cost for the AES ω and unit battery maintenance cost for the AES ω respectively. The third part $\sum_{\omega\in\Omega} C_{tran,\omega}\sum_{(m,u)\in Z,m\neq u}\zeta_{\omega,mu}^t\Delta T$ indicates the cost of AES transportation which calculates the AES trace among ports. $\zeta_{\omega,mu}^t$ is binary variables, 1 if AES ω is on arc (m, u) in time span t, 0 otherwise. The last term $\sum_{\omega\in\Omega} C_{bat,\omega}\sum_{m\in M}|P_{ch,\omega m}^t + P_{dch,\omega m}^t|\Delta T$ provides the shipborne power storage system cost, namely battery maintenance cost. $P_{ch,\omega m}^t$ / $P_{dch,\omega m}^t$is charging/discharging power of AES ω from/to port m in time span t.

In order to effectively dispatch the power resources, the travel of the AES is modeled by using the temporal-spatial network model (TSN model), which describes the vessels' path in 24 h and the exchange of power with each port in detail.

3 THE MODEL OF AES-BASED DISPATCHING

3.1 Temporal-Spatial Constraints of AES

In the TSN model, the transportation time of AESs among nodes is a time span. In practice, it makes hard to be consistent for each transit between ports (nodes). Therefore the virtual port (virtual node) is introduced, i.e. node 4 is introduced in between node 2

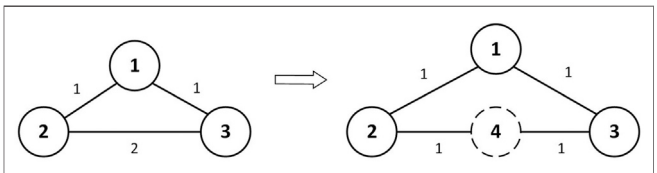

FIGURE 2 | TSN model.

and node 3, so that the travel time of each arc between nodes (including virtual nodes) is exactly a time span, as shown in **Figure 2** (Yao et al., 2018). The virtual node is similar to others except that it cannot be used as dock to transfer power. The TSN model can completely represent the whole process of AESs' transportation and transmission.

The TSN-based AES model is formulated as follows:

$$\sum_{(m,n)\in Z} \zeta^t_{\omega,mu} = 1 \tag{2}$$

$$\sum_{(m,n)\in Z^-_m} \zeta^t_{\omega,mu} = \sum_{(m,n)\in Z^+_m} \zeta^{t+1}_{\omega,mu}, \forall m \in M \cup M_v \tag{3}$$

$$\sum_{(m,n)\in Z^+_m} \zeta^1_{\omega,mu} = \zeta^0_{\omega,m}, \forall m \in M \cup M_v \tag{4}$$

$$\zeta^t_{\omega,mu} + \zeta^{t+1}_{\omega,um} \le 1, (m,n) \in Z, m \ne u \tag{5}$$

$$(\forall \omega = \Omega t \in T)$$

Eq. 2 ensures that there is only one state for each AES at the same time, travel or berth, and the range is given. **Eq. 3** ensures the flow of AES in each port that the entering and exiting number of AESs at any node is equal in sequential time. The constrain holds true for the virtual node. Particularly, constrain **Eq. 4** is a special case of constrain **Eq. 3** which express at the initial time. The constrain **Eq. 5** place restrictions on AESs' trace that AESs cannot course reversal immediately, which means it is necessary to *via* the intermediate port or berth a time space at the target port u. The constrain provides possibility for AESs to berth for power transmission and avoids unnecessary frequent round trip. The above-mentioned constrains do work for AESs in each time span.

According to the TSN model being in line with discrete time span in planning horizon of DGs, TSN model is suitable for scheduling in both representing the location of the AES and ensuring the AES mobile trajectory. Meanwhile, the AES and port microgrid operation are also oblige to guarantee.

3.2 Operation Constraints of AESs

When AESs berth at any port for power resource exchange, the following constraints should be met:

$$P^t_\omega + P^t_{\mathrm{B},\omega} + \zeta^t_{\omega,mm}\left(P^t_{ch,\omega m} - P^t_{dch,\omega m}\right) = P^t_{SE} + \left(1 - \zeta^t_{\omega,mm}\right)P^t_{PL} \tag{6}$$

$$0 \le P^t_{ch,\omega m} \le \zeta^t_{\omega,mm} P^{\max}_{ch,\omega m} \tag{7}$$

$$0 \le P^t_{dch,\omega m} \le \zeta^t_{\omega,mm} P^{\max}_{dch,\omega m} \tag{8}$$

$$0 \le \sum_{m\in M} P^t_{ch,\omega m} \le I^t_{ch,\omega} P^{\max}_{ch,\omega} \tag{9}$$

$$0 \le \sum_{m\in M} P^t_{dch,\omega m} \le I^t_{dch,\omega} P^{\max}_{dch,\omega} \tag{10}$$

$$0 \le P^t_\omega \le P^{\max}_\omega \tag{11}$$

$$SOC_{\min} \le SOC^t_\omega \le SOC_{\max} \tag{12}$$

$$I^t_{ch,\omega} + I^t_{dch,\omega} \le \sum_{m\in M} \zeta^t_{\omega,mm} \tag{13}$$

$$(\omega \in \Omega, m \in M, t \in T)$$

Eq. 6 shows the balance of supply and demand of AESs when berthing for power exchange. P^t_ω, $P^t_{\mathrm{B},\omega}$ are power of diesel generator in AES ω in time span t and power of battery in AES ω in time span t respectively, part of supplyment. P^t_{SE} and P^t_{PL} are service and propulsion loads of ω-th AES in time span t. The constraint ensures that AESs can plunge into joint restoration scheme after meeting their own basic power demand, which averts the damage of shipborne equipment and underpower. $\zeta^t_{\omega,mm} = 1$ indicates AESs berth at port m in time span t. Meanwhile, $1 - \zeta^t_{\omega,mm} = 0$ cause $(1 - \zeta^t_{\omega,mm})P^t_{PL} = 0$ which presents propulsion is no longer needed while AESs berth at port.

Constraint **Eqs. 7–12** represent the bound of the ship borne equipment to maintain their normal operation individually. Constraints **Eqs. 7**, **8** establish the feasible set of charging/discharging power of AES ω to port m. $\zeta^t_{\omega,mm}$ restricts charging/discharging of AESs after berthing. Constraints **Eqs. 9**, **10** define the charge/discharge power constraints associated with AESs' operation mode. $I^t_{ch,\omega}$ / $I^t_{dch,\omega}$, binary variables, represents charging/discharging state of AES ω in time span t restricting by $\zeta^t_{\omega,mm}$. Taking DG into account, the power of shipborne DG is limited, and its bound is shown as constraint **Eq. 11**. In addition, the feasible range of single charge/discharge of shipboard power storage system, i.e. battery, are expressed in **Eq. 12**. The constraint **Eq. 13** is that the AES either transfers power with the port microgrids or be idle while it stops at the port. However, charging and discharging of each AES cannot cooperate simultaneously.

With regard to AESs' trajectory and operation constrains are expressed on above. The constraint set of port microgrid are described as next section.

3.3 Operation Constrains of Port Microgrid

The port is regarded as a complete microgrid from the grids' perspective, which is a small power distribution system applied with equivalent distributed generation, energy storage system, local load and protection device connected by DC bus and controlled by grid control centre. While AESs are connected, the working constraints of microgrid are described as follows:

$$PD^t_{DG,m} + PD^t_{r,m} + P^t_{R,m} = P^t_{DG,m} - \sum_{\omega\in\Omega} P^t_{ch,\omega m} + \sum_{\omega\in\Omega} P^t_{dch,\omega m} \tag{14}$$

$$0 \le PD^t_{r,m} \le PD^t_m \tag{15}$$

$$0 \le P^t_{DG,m} \le P^{\max}_{DG,m} \tag{16}$$

$$E^{t+1}_{DG,m} = E^t_{DG,m} - P^{t+1}_{DG,m}\Delta T \tag{17}$$

$$E^{\min}_{DG,m} \le E^t_{DG,m} \le E^{\max}_{DG,m} \tag{18}$$

$$(\forall m \in M, t \in T)$$

Eq. 14 stands for the co-ordination of supply and demand of the microgrid when AESs are connected and carry out charging/discharging operation. $PD^t_{DG,m}$, $PD^t_{r,m}$ and $P^t_{R,m}$ are load of DG in port m in time span t, load restored at port m in time span t and power reserve of MG in port m in time span t, respectively. As the result of actual load restriction, restored load should be bound which is shown in constrain **Eq. 15**. Constrain **Eq. 16** establish the feasible set of the equivalent distributed generation power in the port. Finally, constrain **Eq. 17** calculates the energy in Port m and constrain **Eq. 18** restrict the capacity of the energy storage system in the microgrid to ensure device working in appropriate range.

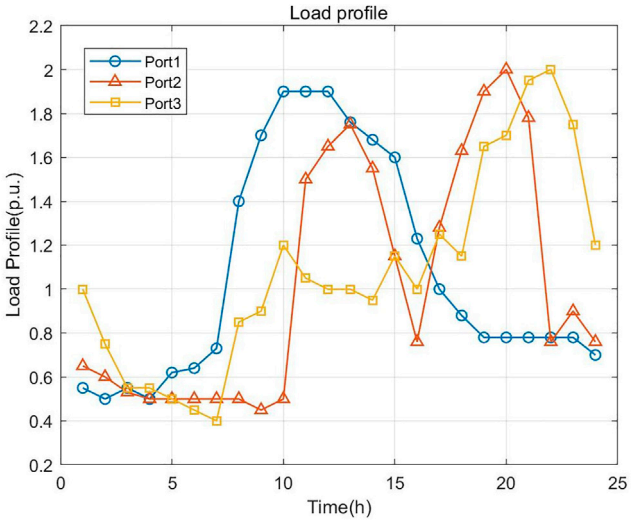

FIGURE 3 | Local load profile.

TABLE 1 | Load interruption cost.

	Port 1	Port 2	Port 3
Cost ($/kWh)	3	10	100

Together with constrains of the AES-based joint post-disaster restoration scheme in Section IV, the dispatching problems pertain to a MILP problem, which can be solved by off-the-shelf business tools effectively, and the constraints are classified into the following 3 categories.

1) Temporal-spatial constraints of TESSs: **Eqs. 2–5**;
2) Operation constraints of AESs: Eq. **Eqs. 6–13**;
3) Operation constraints of port microgrid: **Eqs. 14–18**.

4 NUMERICAL RESULTS

In this section, the case studies are performed on the test system composed of three ports and four AESs to verify the validity of the proposed AES-based joint scheduling scheme. The optimization model is implemented using MATLAB 2020A, on a computer equipped with Intel Core i7 3.4 GHz CPU and 8 GB RAM.

All test problems are solved by intlinprog function in the commercial software MATLAB. "Intlinprog" is a mixed-integer linear programming solver, the abbreviation of mixed-integer linear programming (MILP).

4.1 Text System

In this study, the local load in adjoin ports at each time follows the load profile as given in **Figure 3**.

Owing to stressing on the difference functions, ports' loads show discrepancy, which leading to load classification on supply side. Study assumes that the load classification of ports from low to high is port 1, port 2 and port 3, which is represented by load interruption cost in **Table 1**, **2**.

TABLE 2 | AES parameters.

AES#	1	2	3	4
Initial position	1	1	2	3
Charging/discharging Power (MW)	0.2			
Energy Capacity (MWh)	1.0			
Initial SOC	0.5			
SOCmax/SOCmin	0.90/0.10			
Charging/discharging Efficiency	0.95/0.95			

TABLE 3 | Economic comparison.

	Case 1	Case 2
Cost ($)	2,625.4	2,535.3

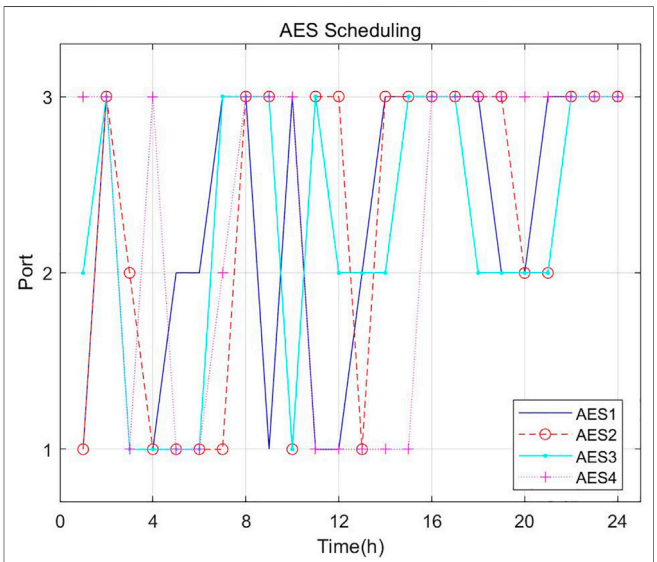

FIGURE 4 | AES scheduling.

Ports, shown in **Figure 1**, are regarded as the microgrid in the island mode, and the local load of each port follows the curve in **Figure 3**. For the sake of simplicity, four AESs with both identical properties and setting out to scheduling simultaneously are considered, which means that all the vessels have exactly the same attributes. The parameters of AESs are shown in **Table 1**. The time horizon is set to 24 h after disaster and time step is 1 h. The unit power generation cost of microgrid is $0.5/kWh. The unit transportation cost of AES is $80 per transit. The unit battery maintenance cost is set to $0.2/kWh.

In the remaining part, two cases are investigated to illustrate the feasibility of joint dispatching of AESs and DSs, as follows.

Case 1) A post-disaster restoration scheme only relies on distributed emergency generator.

Case 2) A joint post-disaster restoration scheme for AESs and generation scheduling in microgrids (MGs) is proposed.

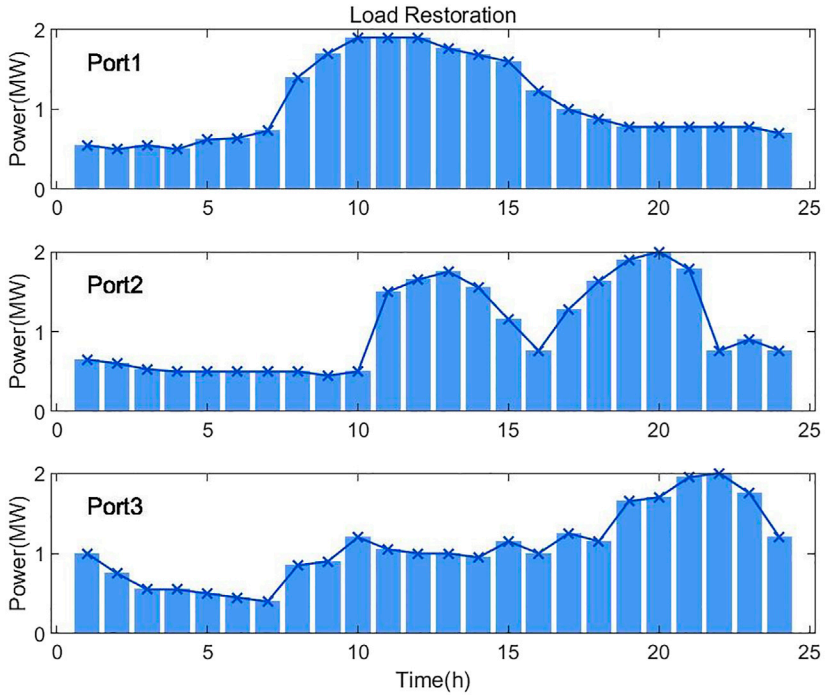

FIGURE 5 | Load restoration in case 2.

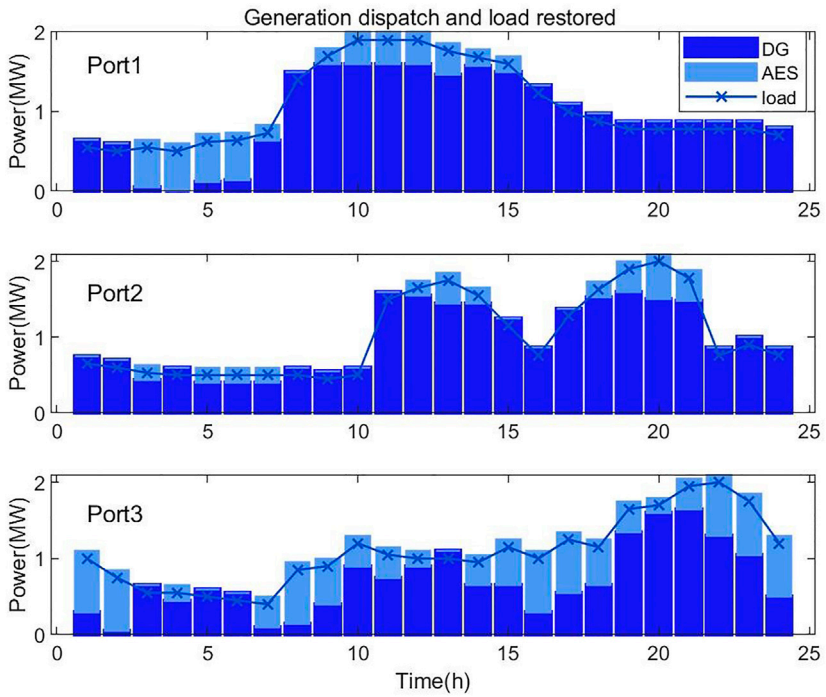

FIGURE 6 | Generation Dispatch and Load Restored in case 2.

4.2 Simulation Results

The comparison of the economic results under two cases is shown in **Table 3**.

Figure 4 provides the dynamics of AES in 24 h. Taking AES3 for instance to illustrate the schedule, it initially berths at port 2 and then through port 3 arrivals at port 1 (01:00–03:00), stays for 3 hours (03:00–06:00) to charge to port 1. Next, AES3 sails to port 3 (06:00–07:00) via port 2, supplies power to port 3 (07:00–09:00). Then departing from port 3 through port 1, port 3 and port 2 in turn (9:00–12:00). Finally moves to port 2 to transit energy to port 2 (12:00–14:00), and then moves to port 3 (14:00–15:00) for discharging 2 hours (15:00–17:00). After that, it returns to port 2 to discharge for 3 hours (18:00–21:00), moves to port 3 (21:00–22:00), starts to supply power from 22:00 to 24:00.

Figure 5 denotes the load recovery of joint post-disaster restoration scheme for AESs and generation scheduling in microgrids. The line diagram is identical as the load curve in **Figure 3**, indicating the change of local load. The variation of the median value of the column chart are perfectly in tune with restoring load. **Figure 6** shows the stacked generation dispatch for microgrids and AESs. The lower part of the bar represents the equivalent active power generated by the distributed emergency equipment; the opposite depicts the feeding of power from AES to the port, which describes the load distribution explicitly. Therefore, it can be observed that the active power within each port is balanced along the optimization horizon while AES participate in restoration.

5 CONCLUSION

As a kind of carrier, AES has the potential to be a distributed emergency generation profit from its immensity power capacity, realizing by the two-way interaction between ship grid and port microgrid through shore power stations. Under the sensible scheduling, it has served as emergency power supply equipment for port microgrid restoration in the post-disaster condition. Therefore, a joint post-disaster restoration scheme for AESs and generation scheduling in microgrids (MGs) has been proposed to minimize the cost of port restoration. The TSN model has been applied to represent AESs in two dimensions: time and space. The port microgrid restoration cost has been established as a mixed-integer linear programming model, which has derived an AES scheduling sequence and generation dispatches for both AESs and resources in port. Taking three adjoining ports and four AESs as examples, the effectiveness of the scheme has been verified.

The comparative simulations have been implemented to illustrate the availabilities of AES on port microgrid restoration. The high-capacity and transportability of AES has guaranteed the feasibility to effectively transfer energy to microgrids, being adjunct to the local loads' operation after disaster, and minimizing losses. The post-disaster island grid can be regarded as a microgrid operating in island mode while the cable disconnection of the inland main grid. Hence, this paper has a certain reference significance for the island post-disaster restoration.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

WX and QS designed the experiments, research methods, and the format analysis. The tools analysis, data processing, and the original draft were carried out by WX. FT performed the writing-review on references. TL contributed to proof reading and project/organisation management. All authors have read and agreed to the published version of the article.

FUNDING

This work is supported in part by the National Natural Science Foundation of China (under Grant Nos. 51939001, 61803064, 61751202, 61903092, 61976033, U1813203); the Science and Technology Innovation Funds of Dalian (under Grant No. 2018J11CY022); the Liaoning Revitalization Talents Program (under Grant Nos. XLYC1908018, XLYC1807046); the Natural Science Foundation of Liaoning (2019-ZD-0151); the Fundamental Research Funds for the Central Universities (under Grant Nos. 3132019345, 3132020103, 3132020125, 3132021140).

REFERENCES

Chanda, S., and Srivastava, A. K. (2016). Defining and Enabling Resiliency of Electric Distribution Systems with Multiple Microgrids. *IEEE Trans. Smart Grid* 7, 2859–2868. doi:10.1109/tsg.2016.2561303

Chen, B., Ye, Z., Chen, C., and Wang, J. (2019). Toward a Milp Modeling Framework for Distribution System Restoration. *IEEE Trans. Power Syst.* 34, 1749–1760. doi:10.1109/tpwrs.2018.2885322

Fang, S., Xu, Y., Wang, H., Shang, C., and Feng, X. (2020). Robust Operation of Shipboard Microgrids with Multiple-Battery Energy Storage System under Navigation Uncertainties. *IEEE Trans. Vehicular Technol.* 69, 10531–10544. doi:10.1109/tvt.2020.3011117

Khodayar, M. E., Wu, L., and Li, Z. (2013). Electric Vehicle Mobility in Transmission-Constrained Hourly Power Generation Scheduling. *IEEE Trans. Smart Grid* 4, 779–788. doi:10.1109/tsg.2012.2230345

Li, H., and Dong, D. (2014). On the Development Status of china's Port Industry and its Role in the Development of National Economy. *Shandong Soc. Sci.*, 144–149. doi:10.14112/j.cnki.37-1053/c.2014.01.004

Lv, R. (2018). Research on Load Guarantee Strategy of Offshore Island Based on mobile Energy Storage Power Supply Network. Ph.D. thesis. Huazhong University of Science and Technology.

Panteli, M., Trakas, D. N., Mancarella, P., and Hatziargyriou, N. D. (2017). Power Systems Resilience Assessment: Hardening and Smart Operational Enhancement Strategies. *Proc. IEEE* 105, 1202–1213. doi:10.1109/jproc.2017.2691357

Parise, G., Parise, L., Martirano, L., Ben Chavdarian, P., Chun-Lien Su, C. L., and Ferrante, A. (2016). Wise Port and Business Energy Management: Port Facilities, Electrical Power Distribution. *IEEE Trans. Ind. Applicat.* 52, 18–24. doi:10.1109/tia.2015.2461176

Ren, J., Chen, J., Jiang, X., Sun, X., Zhang, S., and Li, W. (2020). Post-disaster Recovery Strategy of Resilient Distribution Network Considering mobile Energy Storage System and Network Reconfiguration. *Electric Power Construction* 41, 86–92. doi:10.3969/j.issn.1000-7229.2020.03.011

Wang, A. (2005). Marine Disaster Losses and Disaster Prevention and Mitigation Strategies in china in Recent Years. *Jiangsu Geology.* 029, 98–101.

Wang, J. (2011). Risk Analysis and Assessment of Typical Natural Disasters in Coastal Ports. Ph.D. thesis. East China Normal University.

Wang, Z., and Wang, J. (2015). Self-Healing Resilient Distribution Systems Based on Sectionalization into Microgrids. *IEEE Trans. Power Syst.* 30, 3139–3149. doi:10.1109/tpwrs.2015.2389753

Yang, D., Qin, J., Pang, Y., and Huang, T. (2021). A Novel Double-Stacked Autoencoder for Power Transformers Dga Signals with Imbalanced Data Structure. *IEEE Trans. Ind. Electron.*, 1. doi:10.1109/TIE.2021.3059543

Yao, S., Peng, W., and Zhao, T. (2018). Transportable Energy Storage for More Resilient Distribution Systems with Multiple Microgrids. *IEEE Trans. Smart Grid* 10, 3331–3341. doi:10.1109/TSG.2018.2824820

Zhang, G., Chu, S., Jin, X., and Zhang, W. (2021). Composite Neural Learning Fault-Tolerant Control for Underactuated Vehicles with Event-Triggered Input. *IEEE Trans. Cybern.* 51, 2327–2338. doi:10.1109/TCYB.2020.3005800

Zhou, B., Xu, D., Li, C., Cao, Y., Chan, K. W., Xu, Y., et al. (2017). Multiobjective Generation Portfolio of Hybrid Energy Generating Station for Mobile Emergency Power Supplies. *IEEE Trans. Smart Grid* 9, 5786–5797. doi:10.1109/TSG.2017.2696982

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Energy Research

METHODS
published: 14 September 2021
doi: 10.3389/fenrg.2021.723319

A Multi-Step Prediction Method for Wind Power Based on Improved TCN to Correct Cumulative Error

Haifeng Luo[1], *Xun Dou*[1]*, *Rong Sun*[2] *and Shengjun Wu*[2]

[1]Nanjing Tech University, Nanjing, China, [2]State Grid Jiangsu Electric Power Co. Ltd. Research Institute, Nanjing, China

Wind power generation is likely to hinder the safe and stable operations of power systems for its irregularity, intermittency, and non-smoothness. Since wind power is continuously connected to power systems, the step length required for predicting wind power is increasingly extended, thereby causing an increasing cumulative error. Correcting the cumulative error to predict wind power in multi-step is an urgent problem that needs to be solved. In this study, a multi-step wind power prediction method was proposed by exploiting improved TCN to correct the cumulative error. First, multi-scale convolution (MSC) and self-attentiveness (SA) were adopted to optimize the problem that a single-scale convolution kernel of TCN is difficult to extract temporal and spatial features at different scales of the input sequence. The MSC-SA-TCN model was built to recognize and extract different features exhibited by the input sequence to improve the accuracy and stability of the single-step prediction of wind power. On that basis, the multi-channel time convolutional network with multiple input and multiple output codec technologies was adopted to build the nonlinear mapping between the output and input of the TCN multi-step prediction. The method improved the problem that a single TCN is difficult to tap the different nonlinear relationships between the multi-step prediction output and the fixed input. The MMED-TCN multi-step wind power prediction model was developed to separate linearity and nonlinearity between input and output to reduce the multi-step prediction error. An experimental comparative analysis was conducted based on the measured data from two wind farms in Shuangzitai, Liaoning, and Keqi, Inner Mongolia. As revealed from the results, the MAE and RMSE of the MMED-TCN-based multi-step prediction model achieved the cumulative mean values of 0.0737 and 0.1018. The MAE and RMSE metrics outperformed those of the VMD-AMS-TCN and MSC-SA-TCN models. It can be seen that the wind power prediction method proposed in this study could improve the feature extraction ability of TCN for input sequences and the ability of mining the mapping relationship between multiple inputs and multiple outputs. The method is superior in terms of the accuracy and stability of wind power prediction.

Keywords: improved TCN, cumulative error, multi-step wind power prediction, self-attentiveness, multi-scale convolution

OPEN ACCESS

Edited by:
Dazhong Ma,
Northeastern University, China

Reviewed by:
Xin Fang,
National Renewable Energy Laboratory (DOE), United States
Zhenzhi Lin,
Zhejiang University, China

***Correspondence:**
Xun Dou
dxnjut@njtech.edu.cn

Specialty section:
This article was submitted to Smart Grids, a section of the journal Frontiers in Energy Research

Received: *10 June 2021*
Accepted: *03 August 2021*
Published: *14 September 2021*

Citation:
Luo H, Dou X, Sun R and Wu S (2021) A Multi-Step Prediction Method for Wind Power Based on Improved TCN to Correct Cumulative Error. Front. Energy Res. 9:723319. doi: 10.3389/fenrg.2021.723319

INTRODUCTION

The increasing depletion of traditional energy sources (e.g., fossil fuels and natural gas) has greatly challenged the development of power systems (Wu et al., 2020). Wind energy will become the most promising clean energy source for its inexhaustible and renewable characteristics (Du et al., 2017). The global installed wind power capacity is expected to reach nearly 800 GW by 2021 (Global Wind Energy Council, 2021). Because wind power output is found to be intermittent and stochastic, an accurate wind power prediction method acts as a vital technical tool to ensure the safe, stable, and economic operation of the power system (Ye and Zhao, 2014). Since wind power is continuously connected to the power system, the requirements for step length of its prediction are gradually increasing and the accuracy requirements are gradually becoming higher. However, the conventional multi-step rolling prediction model should exploit the wind power predicted at the previous moment to predict the wind power at the subsequent moment and the prediction result at the subsequent moment will accumulate the prediction error of the previous moment. The cumulative error of wind power will increase as the number of prediction steps rises continuously (Chen et al., 2017). The phenomenon will increase the difficulty of the multi-step prediction of wind power. Accordingly, correcting the cumulative error to conduct the multi-step prediction of wind power should be solved urgently.

On the whole, the existing multi-step prediction of wind power has been conducted based on the single-step prediction. Relevant researchers have adopted a range of methods for the single-step and multi-step predictions of wind power. The mentioned methods can fall into three main categories, i.e., statistical methods, physical methods (Wu et al., 2017), and combined prediction methods (Han et al., 2019). The physical method refers to a wind power prediction method based solely on the historical wind power data and the Numerical Weather Prediction (NWP) data (Louka et al., 2008; De Giorgi et al., 2011; Cassola and Burlando, 2012; Liu et al., 2020a). The physical methods generally apply to the single-step prediction. The single-step wind power prediction based on the physical method exhibits the advantages as follows: the small amount of data relied on, the simplicity of the model, and the convenience and speed of prediction. However, large errors exist between the NWP data and real weather data, and spatial differences are identified between NWP data and wind farms, thereby causing the accuracy of the physical method for the single-step wind power prediction to be generally low. The statistical method refers to a wind power prediction method, updating and adjusting the model weights and parameters based on the error between the predicted and true values. It usually falls into probabilistic statistical models, machine learning models, and deep learning models. Probabilistic statistical models consist of Auto-Regressive Moving Average Model (ARMA) (Wang et al., 2015a) and Autoregressive Integrated Moving Average Model (ARIMA) (Cao et al., 2019; Liu et al., 2020b). It can more effectively follow the wind power forecast than physical methods. Physical methods can better follow the trend of wind power, whereas a large get error occurs in the multi-step prediction. Machine learning models consist of Support Vector Machine (SVM) (Zhang et al., 2016), Random Forest (RF) (Liu et al., 2015), and Hidden Markov Model (HMM) (Lahouar and Ben Hadj Slama, 2017), all of which exhibit better single-step prediction accuracy than that of conventional probabilistic models. Deep learning models comprise Neural Network Model (Zhou et al., 2018), Long Short-Term Memory (LSTM) (Li et al., 2018; Li et al., 2020; Liu and Liu, 2021), and Gated Recurrent Unit (GRU) (Hochreiter and Schmidhuber, 1997; Chung et al., 2014). They are capable of fully exploiting the temporal and spatial characteristics of the input sequence to improve further the accuracy of the single-step prediction of wind power. Statistical methods can effectively extract the linear and nonlinear relationships of historical data and promote the single-step prediction of wind power to be more accurate, whereas it is difficult to conduct an accurate and stable multi-step prediction of wind power series with strong randomness and obvious noise signals for their single structure and high data quality requirements (Tascikaraoglu and Uzunoglu, 2014; Wu et al., 2019).

The combined prediction method refers to a wind power prediction method that maintains the advantages of all single prediction models to achieve more accurate and stable predictions. It is generally used in wind power multi-step prediction. The literature (Lin and Liu, 2011; Wang et al., 2015b; Wang et al., 2020a) has combined VMD (Liu et al., 2018) and GRU to form a combined model for the multi-step prediction of wind power. Moreover, compared with a single model, the combined model single-step prediction results can better track the variation of wind power; however, it is limited by the VMD model and the cumulative error occurs in the multi-step prediction. The literature (Catalao et al., 2010) has employed a hybrid prediction model with wavelet transform, a particle swarm algorithm, and an integrated adaptive network fuzzy inference system. Such a model achieved better mean absolute percentage error (MAPE) and normalized mean absolute error (NMAE) of the single-step prediction than those of the single model; however, the identical problem of cumulative error existed. The literature (Wang et al., 2020b) has combined LSTM, RF, VMD, and wavelet transform (WT) to build a multi-timescale wind power prediction model, and the combined prediction model outperformed others in multiple timescales. The above-mentioned combined prediction method has better prediction performance compared with physical and statistical methods, while no corresponding solution has been given for the multi-step prediction cumulative error.

Combined with the above-mentioned methods, given the cumulative error of existing studies in the wind power multi-step prediction and the incomplete input sequence feature extraction of TCN application in the wind power prediction, this study proposed a wind power multi-step prediction method based on improved TCN to correct the cumulative error. The main contributions of this study are as follows:

1) The MSC-SA-TCN model was established to reduce difficulty in extracting the temporal and spatial features of different

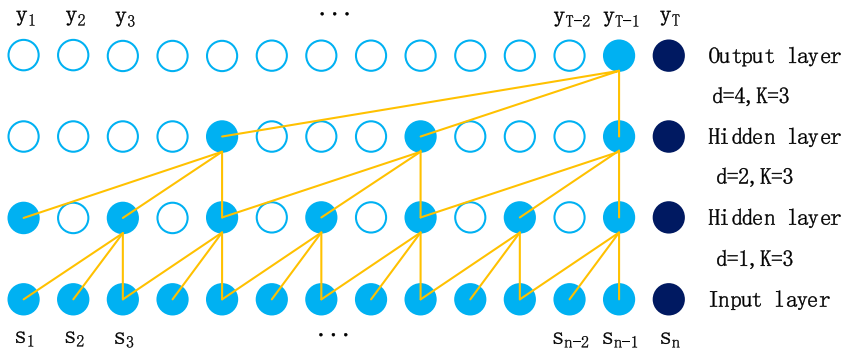

FIGURE 1 | Causal convolution and null convolution in TCN.

scales of the input sequence with the single-scale convolutional kernel of TCN using MSC and SA. It can recognize and extract different features of the input sequence to promote the single-step prediction of wind power to be more accurate and stable.

2) The MMED-TCN multi-step wind power prediction model was proposed to develop the mapping relationship between the output and input of TCN multi-step prediction using the codec of multi-channel time convolutional network with multiple inputs and multiple outputs. Such an effort aims to improve the problem that the different nonlinear relationships between multi-step prediction outputs and fixed inputs are difficult to mine. It can separate linearity and nonlinearity between the input and output to reduce the cumulative error of the multi-step rolling wind power prediction.

This study is organized as follows. In *Design of Multi-Step Prediction Model Based on Improved TCN*, the design of the improved TCN multi-step prediction model is elucidated. In *Algorithm Flow*, the algorithmic process of the wind power multi-step prediction based on improved TCN for correcting the cumulative error is illustrated. In *Experiment and Analysis*, LSTM, VMD-AMS-TCN, and other models are adopted to compare the experiments with the MSC-SA-TCN model and MMED-TCN model proposed in this study. The experimental results are analyzed specifically to verify the superiority of the model proposed in this study. In *Conclusion*, relevant conclusions and subsequent research directions are given.

DESIGN OF MULTI-STEP PREDICTION MODEL BASED ON IMPROVED TCN

Introduction to the TCN Model

On the whole, the TCN model consists of causal convolution, dilated convolution, and residual block. It has a more lightweight network structure than CNN, LSTM, and GRU (Drdgomiretskiy and Zosso, 2013; Bai and Koltun, 2018). The perceptual field of the network can be altered according to the filter size. It is more conducive to the prediction of time series.

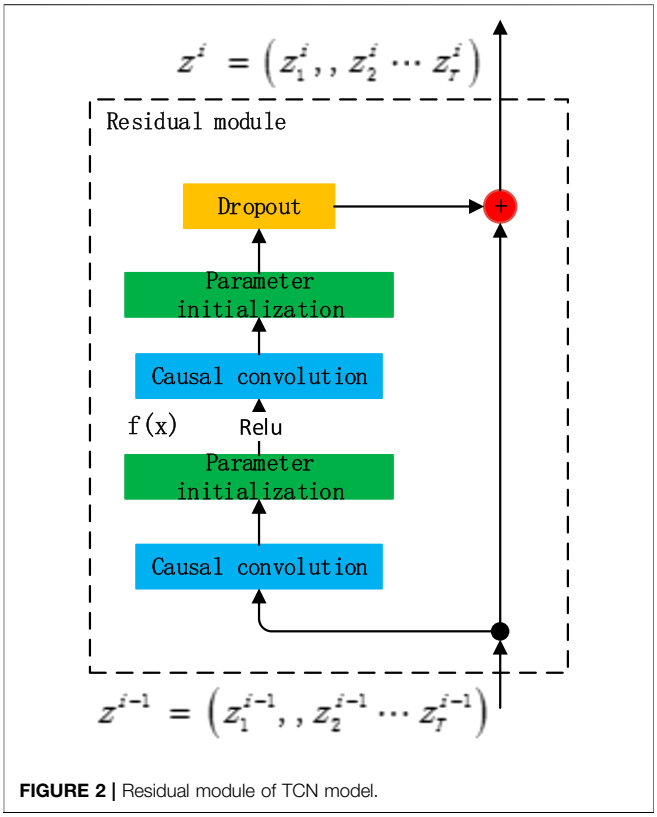

FIGURE 2 | Residual module of TCN model.

Set the filter $F=(f_1,f_2,...,f_K)$, the number is F_n, and the output sequence information is $Y=(y_1,y_2,...,y_s)$, where the input is $S=(s_1,s_2,...,s_n)$, where $s_i i\in[1,n]$ is the column vector. The following equation gives the causal null convolution of s_t at moment t:

$$F(s_t) = (S*_d F)(s_t) = \sum_{k=1}^{K} f_k \cdot s_{t-d(K-k)}, \quad (1)$$

where d denotes the expansion factor, K denotes the filter size, and the formula for the perceptual field is $RF=(K-1)d+1$. The causal and null convolution for $K = 3$ in TCN are given in **Figure 1**.

The TCN model introduced the residual module. The problem of gradient explosion and network degradation in deep

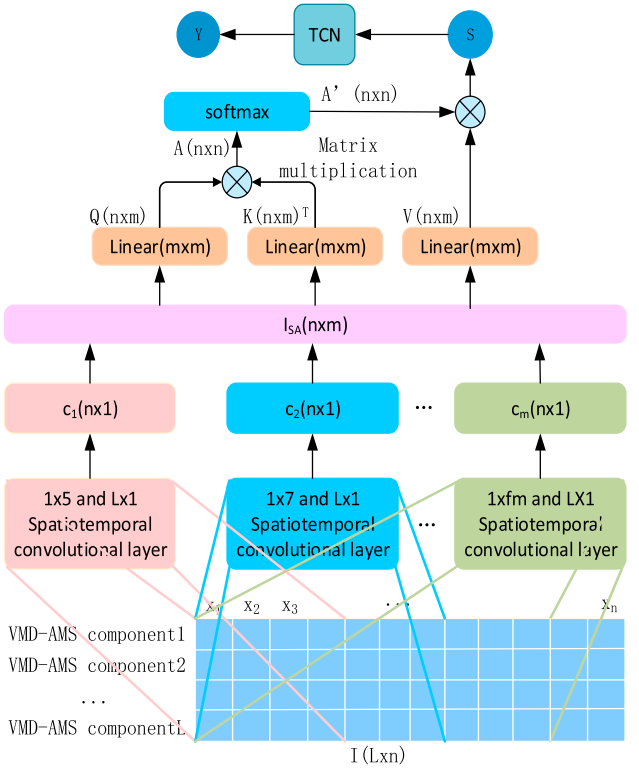

FIGURE 3 | Structure of MSC-SA-TCN model.

conventional neural networks was solved. The residual module of the TCN model is given in **Figure 2**.

In the figure, z^{i-1} is the input of the network at layer *i-1* and z^1 is the output of the network at layer *i*. The calculation in the figure can be expressed as follows:

$$f(x) = W_2\sigma\left(W_1 z^{i-1} + b_1\right) + b_2, \tag{2}$$

$$z^i = f(x) + z^{i-1} \tag{3}$$

where W_1, W_2, b_1, and b_2 denote the mapping parameters to be learned by the TCN; $\sigma(\cdot)$ is the Rule function.

According to the mentioned brief and literature research, the current TCN faces difficulty in extracting multi-scale temporal and spatial features of input sequences and in mining the different nonlinear mapping relationships between multi-step prediction outputs and fixed inputs.

MSCSA-TCN Model

To solve the problem that the size of the convolution kernel of the conventional TCN model is fixed, in order to reduce the difficulty in extracting the multi-scale temporal and spatial features extracted from the input sequence, this study proposed an improved TCN model based on MSC-SA. First, different scales of convolution were adopted to extract the complete time-space features of wind power sequences. Subsequently, the self-attentive mechanism was used to mine the correlation among the features and distinguish the important features from the non-important ones. Lastly, the output of MSC-SA acted as the input of TCN. The structure of the MSC-SA-TCN model is illustrated in **Figure 3**.

c_i is the output of MSC, expressed as follows:

$$c_i = K^i_{L\times 1} * \left[I_{1\times k_i} * K^{ij}_{1*k_i}\right] \quad j \in [1, n], \tag{4}$$

where * is the convolution operation; k_i is the layer *i* convolution kernel scale; *L* represents the number of input features; I_{1xki} is the output of the VMD-AMS module; $K^{i,n}_{1\times k_i}$ denotes the *j*th temporal convolution kernel in layer *i* with size $1xk_i$; $K^i_{L\times 1}$ is the spatial convolution kernel in layer *i* with size *Lx1*.

Q,K,V of the attention mechanism part of the figure can be expressed as follows:

$$\begin{aligned} Q &= W^Q_{n\times n} I_{SA} + b^Q_{n\times m}, \\ K &= \left(W^K_{n\times n} I_{SA} + b^K_{n\times m}\right)^T, \\ V &= W^V_{n\times n} I_{SA} + b^V_{n\times m}, \end{aligned} \tag{5}$$

where W^* and b^* represent the weight matrix and bias matrix; then, the output of the MSC-SA module can be derived as follows:

$$S = soft\max(QK)V. \tag{6}$$

MMED-TCN Model

For the cumulative error in the wind power multi-step prediction by traditional single TCN, an improved TCN model based on multiple outputs was proposed in this study. The specific structure of the model is shown in **Figure 4**. The model reduced the cumulative error in the wind power multi-step prediction using multi-channel TCN to extract the nonlinear mapping relationship between input and output of different prediction steps.

The input of the multi-output TCN model is the output *S* of MSC-SA with dimension *kxm*, and its individual output is expressed as T_{Fnx1}. *Fn* is the number of TCN filters, and the final output was obtained after linear transformation and Relu function, which can be expressed as follows:

$$y_i = \mathrm{Re}\,lu\left(W_{1\times Fn,i} T_{Fn\times 1,i} + b_i\right) \quad i \in [1, s], \tag{7}$$

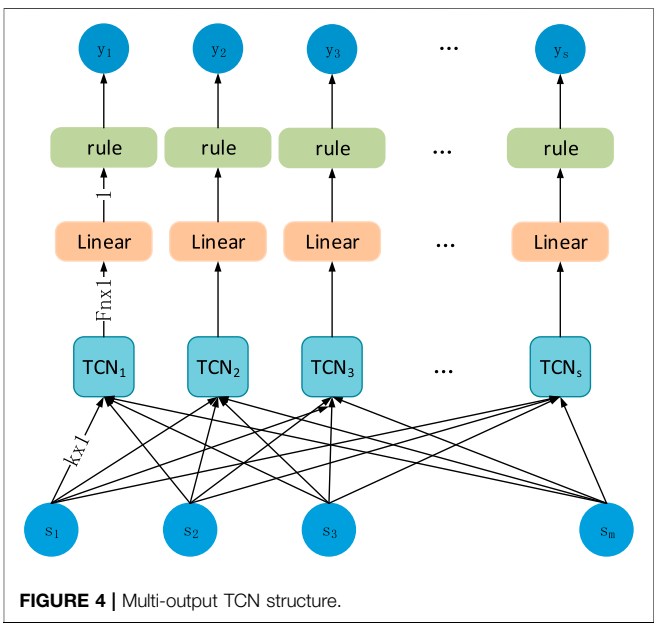

FIGURE 4 | Multi-output TCN structure.

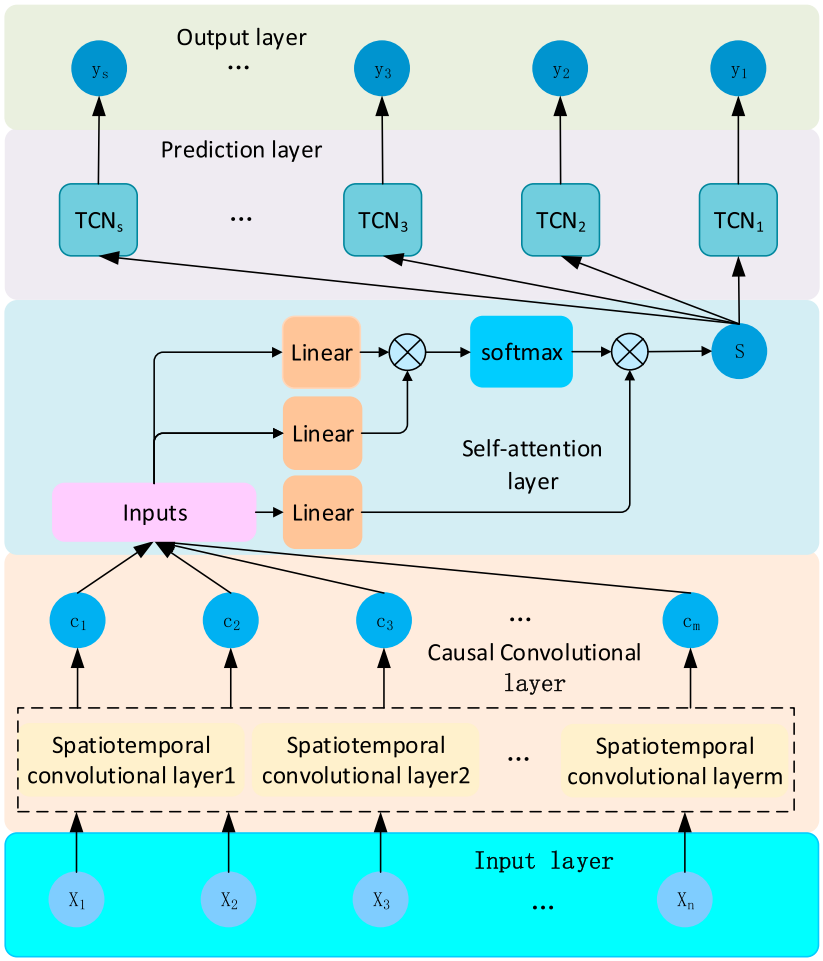

FIGURE 5 | MMED-TCN model structure.

where $W_{Fnx1,i}$, b_i are the weights and biases of the predicted output at step i, respectively.

The MMED-TCN model was obtained using the multi-output TCN model as the decoding layer and using the MSC-SA model as the coding layer. The structure of the MMED-TCN model is given in **Figure 5**.

ALGORITHM FLOW

Based on the design of a multi-step prediction model with improved TCN, the algorithm flow of wind power multi-step prediction based on improved TCN to correct cumulative error proposed in this study is illustrated in **Figure 6**. The specific steps are elucidated as follows:

1) First, the input raw wind power sequence P was preprocessed, mainly including outlier processing (negative value, exceeding full power value and garbled code) and vacant value filling, and the wind power sequence after processing is expressed as P'.
2) Then, P' was decomposed using VMD to obtain the principal component sequence $IMFs = \{IMF_1, IMF_2, \ldots, IMF_{L-1}\}$ and the noise sequence E. The IMFL was obtained by smoothing E using AMS.
3) Next, the MSC-SA model was used to extract the temporal and spatial features of the IMFs series to obtain the output S.
4) Finally, the final prediction model was selected according to the number of prediction steps. The MSC-SA-TCN model was selected for the single-step prediction. The MMED-TCN model was selected for the multi-step prediction. The prediction results were evaluated.

EXPERIMENT AND ANALYSIS

Data Source

The wind power data used in this study were obtained from the actual measurement data of the wind farms in Shuangzitai, Liaoning, and Kqi, Inner Mongolia, from December 2019 to October 2020, with a sampling interval of 15 min. Nearly 28,000 data pieces were available for each wind farm, and the first 10,000 were taken as the experimental data in this study, the first 80% of which acted as the training set and the last 20% as the test set.

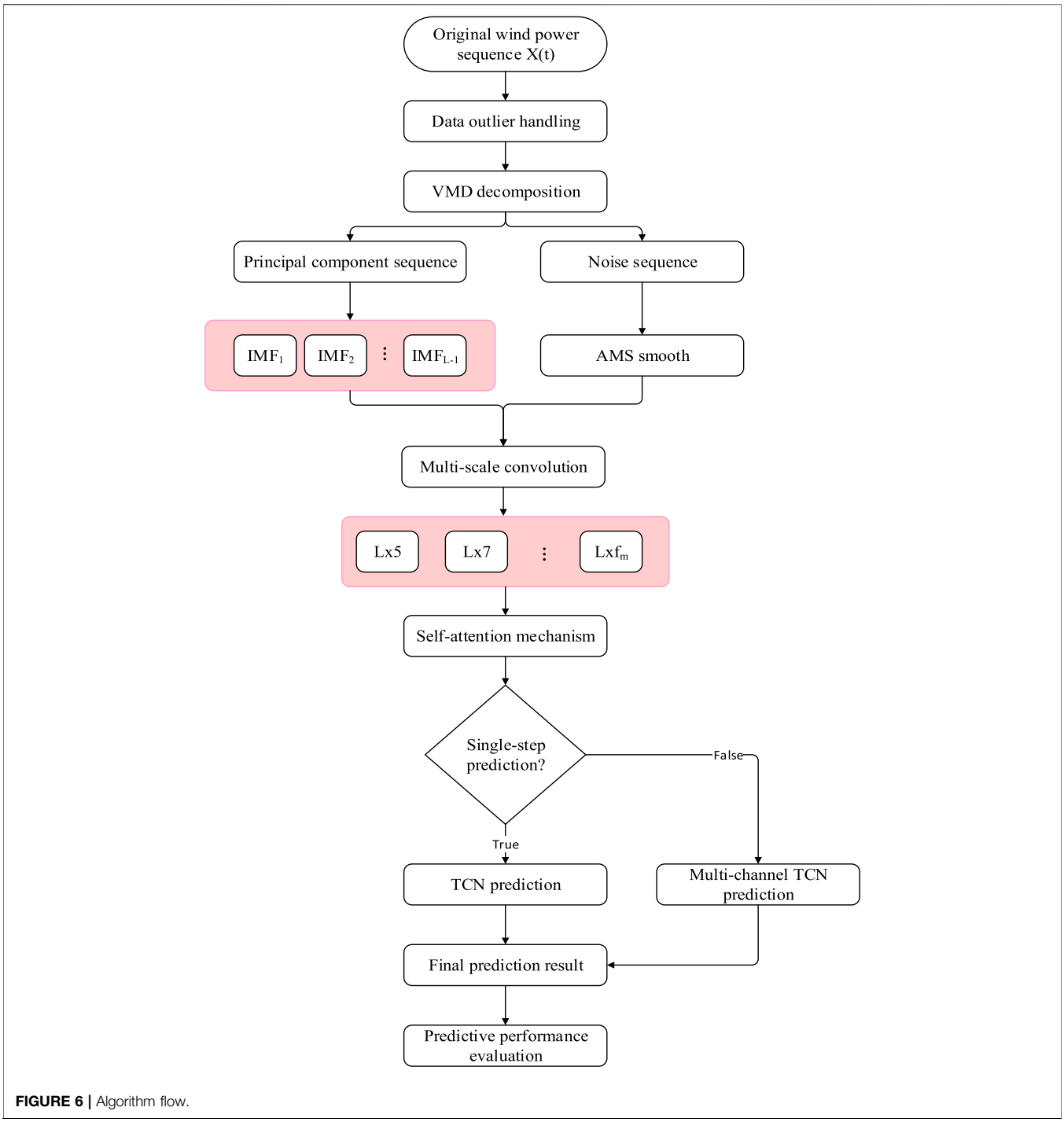

FIGURE 6 | Algorithm flow.

Data Processing

The data employed in this study have negative, overfull, garbled, and vacant values. In this study, the negative values were processed by directly setting zero, and the wind power values before and after the two moments were used to correct or fill the overfull, garbled, and vacant values. Since the installed capacity of each wind power station was different, thereby causing the difference of its power generation, the experimental data were normalized to better assess the experimental results using the following:

$$P_i'' = \frac{P_i' - \min(P')}{\max(P') - \min(P')}. \tag{8}$$

Evaluation Metrics

In this study, six metrics were adopted to assess the performance of the model, i.e., Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), MAE lift (P_{MAE}), and RMSE lift (P_{RMSE}) of the prediction results of the two models and MAE

TABLE 1 | Experimental comparison model and main parameters.

Number	Comparison model	Main parameters			
1	ARIMA	LSTM	TCN	ARIMA	VMD
2	LSTM	1) Number of nodes in the input layer: 200. 2) Number of convolution kernels: 10, size: 3. 3) Residual layer: 2. 4) Dilated convolution: Ye and Zhao (2014); Liu et al. (2015); Du et al. (2017); Liu et al. (2020a); Wu et al. (2020). 5) Training batch size: 32. 6) Maximum number of training iterations: 100.	1) Number of nodes in the input layer: 200. 2) Number of implied layers: 4. 3) Number of nodes in the output layer: 1. 4) Training batch size: 32. 5) Maximum number of training iterations: 100.	1) Autocorrelation order: 24. 2) Moving average order: 4. 3) Difference order: 1.	1) Number of decompositions: 20. 2) Penalty factor: 1,000.
3	TCN				
4	EMD-TCN				
5	VMD-AMS-TCN				

accumulation ($Adde_{MAE,ij}$) and RMSE accumulation ($Adde_{RMSE,ij}$) from i-step prediction to j-step prediction. The specific equations are expressed as follows:

$$MAE = \frac{1}{n}\sum_{k=1}^{n}\left|y_i - \hat{y}_i\right|, \tag{9}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2}, \tag{10}$$

$$P_{MAE} = \frac{Model2_{MAE} - Model1_{MAE}}{Model2_{MAE}}, \tag{11}$$

$$P_{MRE} = \frac{Model2_{RMSE} - Model1_{RMSE}}{Mode12_{RMSE}}, \tag{12}$$

$$Adde_{MAE,ij} = MAE_i - MAE_j \quad i > j, \tag{13}$$

$$Adde_{RMSE,ij} = RMSE_i - RMSE_j \quad i > j, \tag{14}$$

where y_i and $\hat{y}_i$ are the true and predicted values at moment i, respectively. $Model_{MAE}$ and $Model_{RMSE}$ represent the MAE and RMSE of the model, respectively, where $Model_1$ is the combined model of this study and $Model_2$ is the comparison model. MAE_i and $RMSE_i$ denote the MAE and RMSE predicted at step i, respectively, and $MAEj$ and $RMSE_j$ are the MAE and RMSE predicted at step j , respectively, where $i, j \in$ (Chen et al., 2017; Du et al., 2017) are the prediction steps.

Experimental Analysis of MSC-SA-TCN Model

Comparison of Experimental Models

To verify the prediction effect of the single-step prediction model MSC-SA-TCN proposed in this study, five models in **Table 1** were used for the experimental comparison in this section, and the parameter optimization of the mentioned models was not the focus of this study. For this reason, the main parameters of each model are listed directly in **Table 1**.

Experiment and Analysis of VMD-AMS Algorithm

The VMD decomposition algorithm was adopted to decompose the processed wind power series P'' into 20 main components *IMFs* and 1 error component E. Subsequently, the AMS algorithm was employed to smooth the tracking of the error component E to determine the 21st *IMF* component. The results of the VMD decomposition and the AMS smoothing are presented in **Figure 7** (the figure presents the results of 3, 6, 9, 12, 15, 18, 21, and so on for different IMFs components). As indicated from this figure, each component processed using the VMD-AMS algorithm was uniformly distributed in the frequency domain, and the reconstruction error was negligible below 0.005.

Experiments and Analysis of the Comparison Model

In this study, three single models (i.e., ARIMA, LSTM, and TCN) were set for the comparative experimental analysis to verify the superiority of the TCN model in the wind power prediction. Next, two sets of hybrid models (i.e., EMD-TCN and VMD-AMS-TCN) were set for the comparative experimental analysis to verify the effectiveness of the VMD-AMS algorithm in improving the prediction accuracy. Experiments were performed for the proposed model MSC-SA-TCN and the other five models, respectively. The experimental results are illustrated in **Figure 8**, and the performance of various models under the evaluation metrics MAE and RMSE is listed in **Table 2**.

As revealed from the comparison of the prediction experiments of the two wind farms in **Table 2**, the MAE of the single models (e.g., ARIMA, LSTM, and TCN) exceeded 1.5, and the RMSE reached over 2.0, aking the trend of wind power difficult to be accurately tracked. The main reason for this phenomenon was that the single prediction model exhibited relatively low sensitivity to the noise components in the wind power series. Among the single models, TCN performs slightly better than the other two single models in MAE and RMSE due to the introduction of causal null convolution. Thus, TCN was endowed with a wider field of perception compared with ARIMA and LSTM, and it was enabled to more effectively obtain the temporal characteristics of the wind power series. Compared with the single model, EMD-TCN and VMD-AMS-TCN models significantly improved their performance in MAE and RMSE. The introduction of EMD and VMD-AMS could effectively separate the noise signals, reduce the effect of noise signals on the prediction accuracy, and increase the prediction accuracy. The prediction performance of the VMD-AMS-TCN model was better than that of the EMD-TCN model. The prediction performance of the MSC-SA-TCN model proposed in this study was better than that of the other models, where the MAE was less than 0.3 and the RMSE was less than 0.4.

To further verify the prediction performance improvement of the proposed MSC-SA-TCN model compared with other models

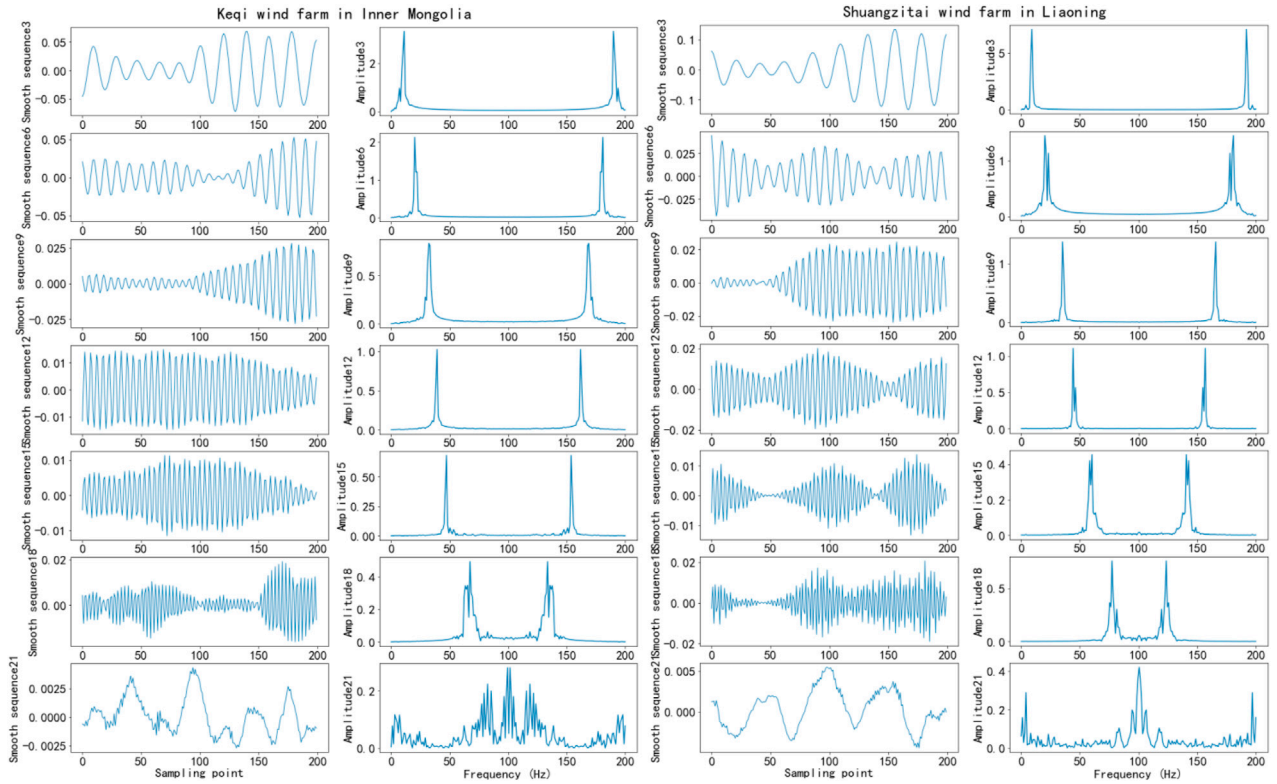

FIGURE 7 | Experimental results of VMD-AMS algorithm.

in this study, the performance improvement of the MSC-SA-TCN model is given in **Table 3**.

According to **Table 3**, the proposed MSC-SA-TCN model in this study improved more than 50% in MAE and more than 40% in RMSE. The average improvement in MAE compared with other models was 80.1%. The average improvement in RMSE reached 80%. The MAE and RMSE predictors of the MSC-SA-TCN model were significantly improved.

To verify the prediction stability of the MSC-SA-TCN model proposed in this study, ten comparison experiments were performed using the VMD-AMS-TCN model and the MSC-SA-TCN model for the Nemengkchi wind farm, and the experimental results are shown in **Figure 9**. The MAE and RMSE of the MSC-SA-TCN model fluctuated from 0.1 to 0.3, while those of the VMD-AMS-TCN model fluctuated from 0.3 to 0.7 and 0.4 to 0.8, respectively. The fluctuation ranges of the MAE and RMSE of the MSC-SA-TCN model were smaller than those of the MAE and RMSE.

Experimental Analysis of Multi-Step Wind Power Prediction

The MMED-TCN wind power multi-step prediction model proposed in this study reduces the cumulative error of the wind power multi-step rolling prediction using the multi-channel TCN technique. To verify the effectiveness of the MMED-TCN model in reducing the cumulative error, the VMD-AMS-TCN and MSC-SA-TCN models were employed to compare the 2-step and 5-step predictions with the MMED-TCN model under two wind farms, where the VMD-AMS-TCN and MSC-SA-TCN models were used for rolling prediction. The experimental results are presented in **Figure 10**, and the performance comparison of the respective model is listed in **Table 4**.

According to **Table 4**, the 2-step and 5-step wind power prediction results of the MMED-TCN model proposed in this study were better than those of the VMD-AMS-TCN and MSC-SA-TCN models overall under the MAE and RMSE evaluation metrics. To verify whether the MMED-TCN model could effectively reduce the cumulative error in the multi-step prediction, the comparison of the cumulative error of each model is listed in **Table 5**.

As indicated from **Table 5**, the MMED-TCN model proposed in this study could effectively reduce the cumulative error in the multi-step prediction, from 2-step prediction to 5-step prediction. Its $Adde_{MAE,ij}$ was less than 0.16, and its mean value was 0.0737, which was better than 0.2127 of the VMD-AMS-TCN model and 0.1991 of the MSC-SA-TCN model. Its $Adde_{RMSE\ ,ij}$ was less than 0.2, and the mean value was 0.1018, which was better than 0.3143 of the VMD-AMS-TCN model and 0.2501 of the MSC-SA-TCN model. To elucidate the effect of the MMED-TCN model in reducing the cumulative error of the multi-step prediction, the histogram of the cumulative error index of each model is presented in **Figure 11**. According to **Figure 11**, the MMED-TCN model outperformed both the VMD-AMS-TCN model and the MSC-SA-TCN model in terms of $Adde_{MAE,ij}$ and $Adde_{RMSE,ij}$. The

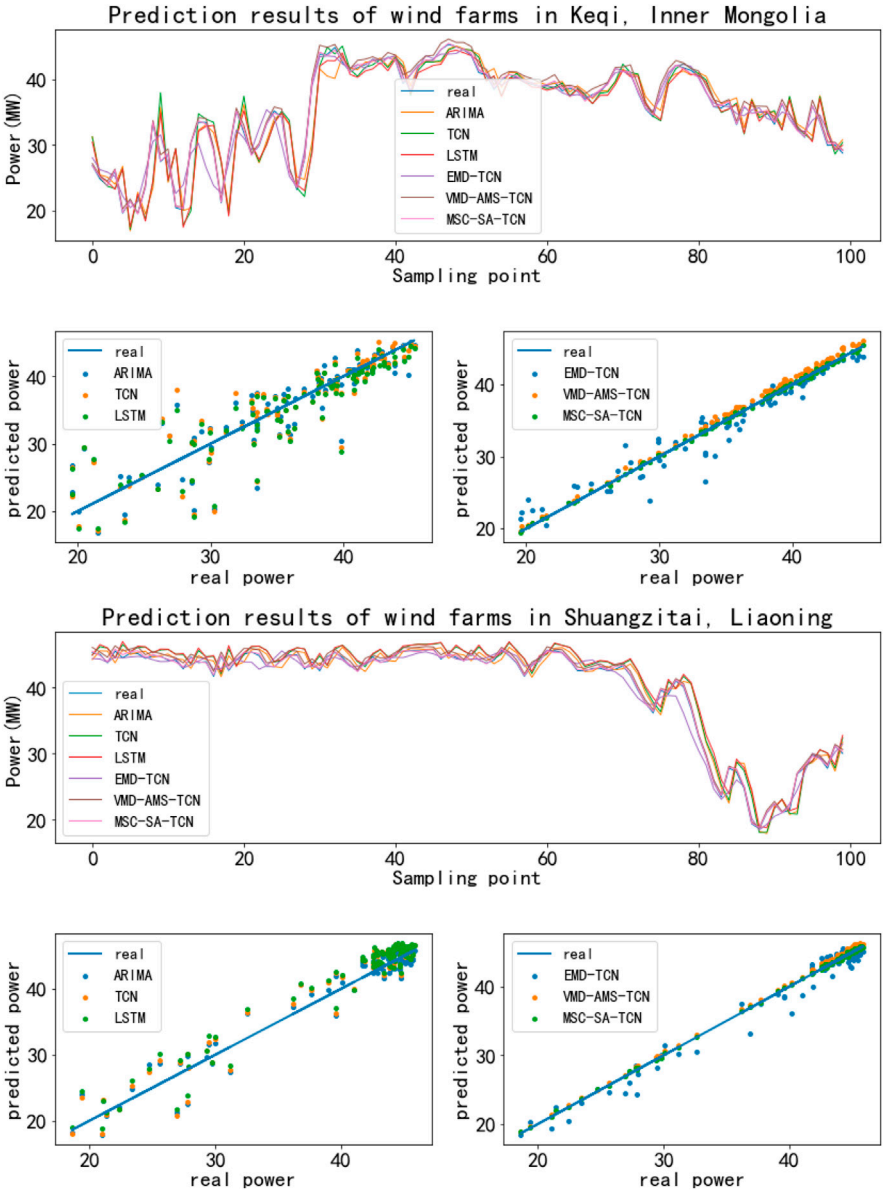

FIGURE 8 | Experimental results of single-step wind power prediction.

TABLE 2 | Comparison of the performance of each prediction model for the two wind farm experiments.

Models	Keqi wind farm in inner Mongolia		Shuangzitai wind farm in Liaoning	
	MAE	RSME	MAE	RMSE
ARIMA	1.7022	2.5257	1.8374	2.6323
LSTM	1.6911	2.4964	1.6749	2.4041
TCN	1.6896	2.3118	1.5803	2.3431
EMD-TCN	0.8664	1.2362	1.1458	1.5513
VMD-AMS-TCN	0.3361	0.4036	0.5454	0.6116
MSC-SA-TCN	0.1330	0.1755	0.2585	0.3445

TABLE 3 | Comparison of the performance improvement of the experimental MSC-SA-TCN model for two wind farms.

Contrast model	Keqi wind farm in inner Mongolia		Shuangzitai wind farm in Liaoning	
	PMAE (%)	PRSME (%)	PMAE (%)	PRSME (%)
ARIMA	92.1	93.1	85.9	86.9
LSTM	93.5	92.9	84.6	85.7
TCN	92.1	92.4	82.0	85.3
EMD-TCN	84.6	85.8	77.4	77.8
VMD-AMS-TCN	60.4	56.5	52.6	43.7

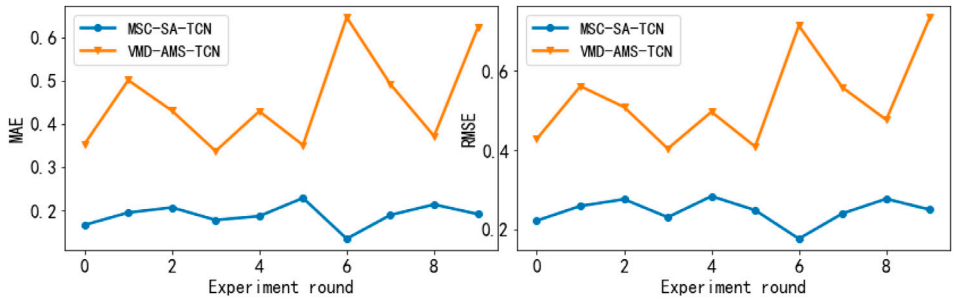

FIGURE 9 | Comparison of prediction stability among models.

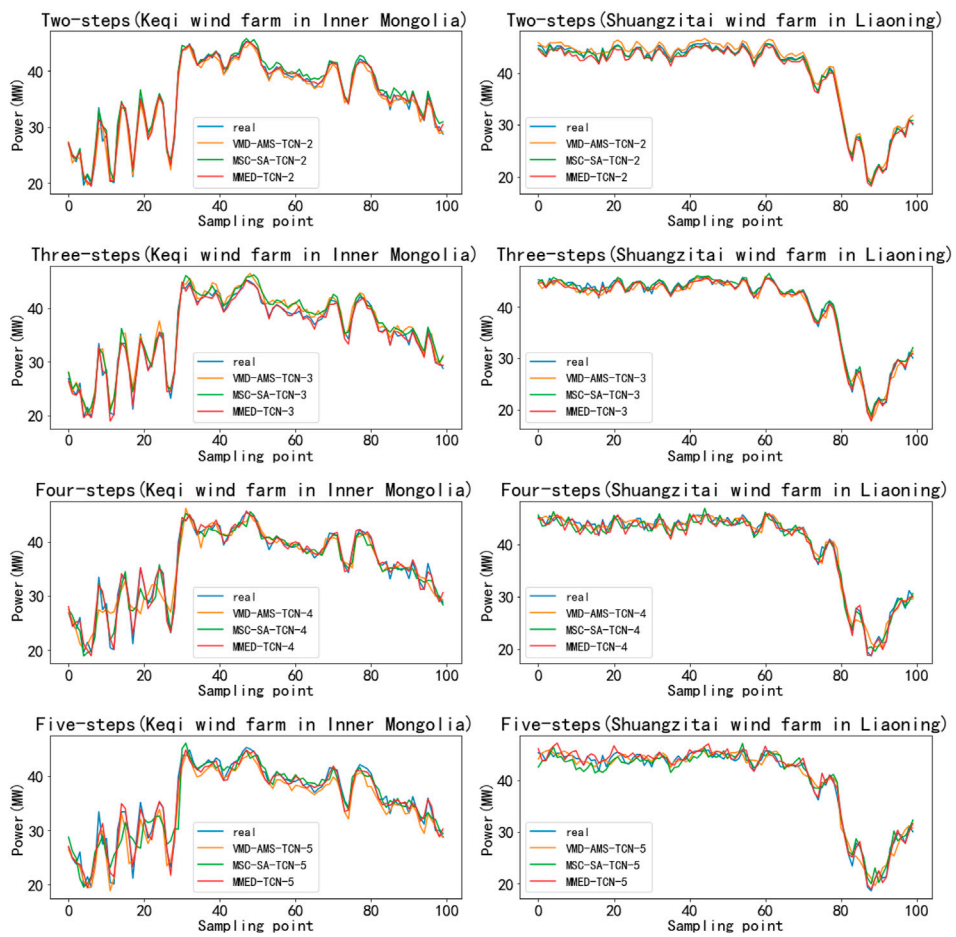

FIGURE 10 | Multi-step wind power prediction experimental results.

TABLE 4 | Comparison of the performance of each prediction model for the two wind farm experiments.

Models	Metrics	Keqi wind farm in inner Mongolia				Shuangzitai wind farm in Liaoning			
		2 steps	3 steps	4 steps	5 steps	2 steps	3 steps	4 steps	5 steps
VMD-AMS-TCN	MAE	0.6363	0.8623	1.0494	1.3022	0.6434	0.8453	1.0665	1.2528
	RMSE	0.8529	1.1338	1.5191	1.8021	0.8618	1.2052	1.5185	1.7985
MSC-SA-TCN	MAE	0.5957	0.6935	0.8766	1.1483	0.4087	0.6876	0.8145	1.0508
	RMSE	0.7645	0.9580	1.2159	1.6104	0.7273	0.9636	1.0865	1.4418
MMED-TCN	MAE	0.3832	0.4075	0.5098	0.6603	0.5586	0.6116	0.6572	0.7237
	RMSE	0.5202	0.5485	0.6925	0.8859	0.7273	0.8198	0.9112	0.9724

TABLE 5 | Comparison of error accumulation of each prediction model for the two wind farm experiments.

Models	Metrics	Keqi wind farm in inner Mongolia				Shuangzitai wind farm in Liaoning				Mean
		I = 2,j = 3	I = 3,j = 4	I = 4,j = 5	Mean1	I = 2,j = 3	I = 3,j = 4	I = 4,j = 5	Mean2	
VMD-AMS-TCN	AddeMAE,ij	0.2260	0.1871	0.2538	0.2223	0.2019	0.2212	0.1863	0.2031	0.2127
	AddeRMSE,ij	0.2809	0.3853	0.2830	0.3164	0.3434	0.3133	0.2800	0.3122	0.3143
MSC-SA-TCN	AddeMAE,ij	0.0978	0.1831	0.2717	0.1842	0.2789	0.1269	0.2363	0.2140	0.1991
	AddeRMSE,ij	0.1935	0.2579	0.3945	0.2820	0.2363	0.1229	0.3553	0.2382	0.2501
MMED-TCN	AddeMAEij	0.0243	0.1023	0.1505	0.0924	0.0530	0.0456	0.0665	0.0550	0.0737
	AddeRMSE,ij	0.0283	0.1440	0.1934	0.1219	0.0925	0.0914	0.0612	0.0817	0.1018

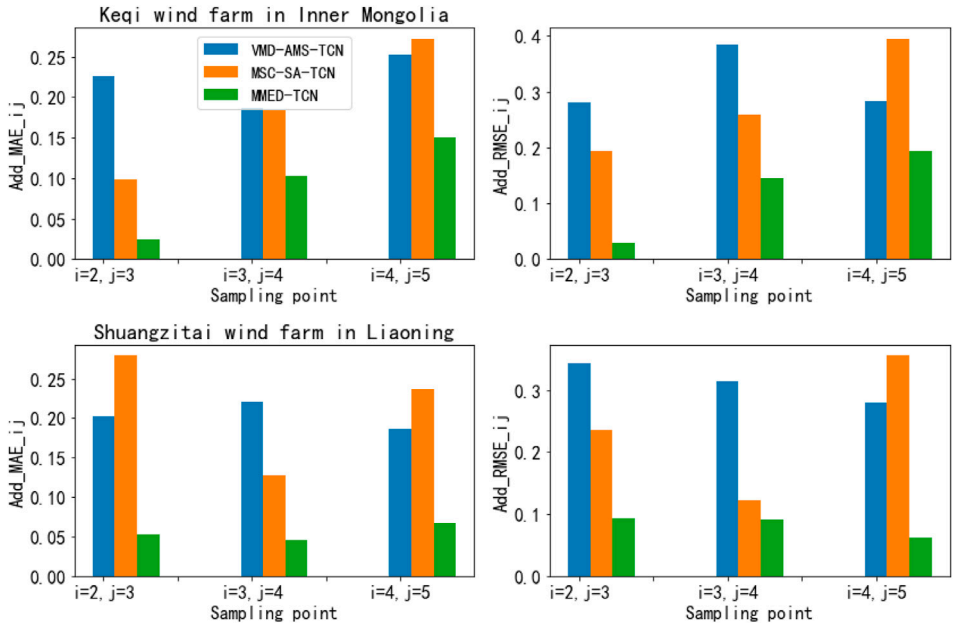

FIGURE 11 | Histogram of error accumulation index of each model.

effectiveness of the MMED-TCN model in reducing the multi-step prediction error was further verified.

CONCLUSION

To cope with the cumulative error in the wind power multi-step prediction, a wind power multi-step prediction method based on improved TCN to correct the cumulative error was proposed in this study. The MMED-TCN multi-step wind power prediction model was built by optimizing the TCN single-scale convolution kernel and single input-output mapping relationship. Based on the model, several experiments were performed on the actual measured data of Liaoning Shuangzitai wind farm and Inner Mongolia Keqi wind farm. The following conclusions could be drawn based on the experimental results.

1) The MSC-SA-TCN model can effectively fix the ability of the TCN's single-scale convolution kernel for input sequence feature extraction and improve the prediction accuracy and stability of

the model by extracting the temporal and spatial features of the input sequence at different scales by MSC-SA. The MAE and the RMSE of the MSC-SA-TCN model decrease by 0.2450 and 0.2476 on average in the single-step prediction, respectively, compared with those of the VMD-AMS-TCN model.

2) The MMED-TCN model is capable of effectively fixing the ability of a single TCN to mine the mapping relationship between multiple inputs and multiple outputs, reducing the effect of the previous step prediction error on the prediction using the multi-channel TCN technique, and effectively reducing the cumulative error of the multi-step rolling prediction. In the multi-step wind power prediction, its MAE cumulative mean value reaches 0.0737, and the RMSE cumulative mean value is 0.1018, better than those of other models.
3) The MMED-TCN model introduced in this study can effectively reduce the cumulative error of the multi-step prediction of wind power. Since the PV power series exhibits similar characteristics with the wind power series, the MMED-TCN model can be adopted to reduce the cumulative error of the multi-part prediction of PV power in the future.

DATA AVAILABILITY STATEMENT

The datasets presented in this article are not readily available because the data comes from the provincial dispatch company, it was non-public data. Requests to access the datasets should be directed to corresponding author XD, dxnjut@njtech.edu.cn. Additional data from the article can be found in the **Supplementary Material**.

AUTHOR CONTRIBUTIONS

HL built the models, analyzed the experiments, and wrote the manuscript. XD provided guidance in writing and revised the manuscript. RS and SW contributed to the writing.

FUNDING

This work was funuded by the State Grid Jiangsu Electric Power Co., Ltd., Research Institute under the Science and Technology Project "Technology Foundation of SGCC (State Grid Jiansu Electric Power Co. Ltd. Research Institute's high-precision power prediction research service based on the spatio-temporal evolution law of new energy power."

ACKNOWLEDGMENTS

The authors acknowledge the State Grid Jiangsu Electric Power Co., Ltd., Research Institute and Nanjing TECH University for their feasibility research support.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fenrg.2021.723319/full#supplementary-material

REFERENCES

Bai, S. J., and Koltun, V. (2018). An Empirical Evaluation of Generic Convolutional and Recurrent Networks for Sequence Modeling. arXiv pre-print server.

Cao, J., Zhou, R., Deng, X., Fan, W., Liu, L., and Sun, J. (2019). Wind Power Prediction Considering the Differential Times of Optimized ARIMA Model[J]. *J. Power Syst. their Automation* 31 (01), 105–111. doi:10.3969/j.issn.1003-8930.2019.01.017

Cassola, F., and Burlando, M. (2012). Wind Speed and Wind Energy Forecast through Kalman Filtering of Numerical Weather Prediction Model Output. *Appl. Energ.* 99 (6), 154–166. doi:10.1016/j.apenergy.2012.03.054

Catalao, J. P. S., Pousinho, H. M. I., and Mendes, V. M. F. (2010). Hybrid Wavelet-PSO-ANFIS Approach for Short-Term Wind Power Forecasting in Portugal. *IEEE Trans. Sustain. Energ.* 2 (1), 1949–3029. doi:10.1109/TSTE.2010.2076359

Chen, N., Xue, Y. S., Ding, J., Chen, Z. L., Wang, W. Z., and Wang, N. B. (2017). Ultra-short-term Wind Speed Prediction Using Spatial Correlation[J]. *Power Syst. Automation* 41 (12), 124–130. doi:10.7500/AEPS20160914002

Chung, J., Gülçehre, Ç., Cho, K., and Bengio, Y. (2014). Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling [J]. *IEICE Trans. Fundamentals Electron. Commun. Comput. Sci.* abs/1412.3555.

De Giorgi, M. G., Ficarella, A., and Tarantino, M. (2011). Error Analysis of Short Term Wind Power Prediction Models. *Appl. Energ.* 88 (4), 1298–1311. doi:10.1016/j.apenergy.2010.10.035

Drdgomiretskiy, K., and Zosso, D. (2013). Variational Mode Decomposition [J]. *IEEE Transactions Signal. Processing* 62 (3), 531–544. doi:10.1109/TSP.2013.2288675

Du, P., Wang, J., Guo, Z., and Yang, W. (2017). Research and Application of a Novel Hybrid Forecasting System Based on Multi-Objective Optimization for Wind Speed Forecasting. *Energ. Convers. Manage.* 150, 90–107. doi:10.1016/j.enconman.2017.07.065

Global Wind Energy Council (2021). *Global Wind Report 2021*.

Han, Z., Jing, Q., Zhang, Y., Bai, R., Guo, K., and Zhang, Y. (2019). A Review of Wind Power Forecasting Methods and New Trends[J]. *Power Syst. Prot. Control.* 47 (24), 178–187. doi:10.19783/j.cnki.pspc.190128

Hochreiter, S., and Schmidhuber, J. (1997). Long Short-Term Memory. *Neural Comput.* 9 (8), 1735–1780. doi:10.1162/neco.1997.9.8.1735

Lahouar, A., and Ben Hadj Slama, J. (2017). Hour-Ahead Wind Power Forecast Based on Random Forests. *Renew. Energ.* 109, 529–541. doi:10.1016/j.renene.2017.03.064

Li, B., Zhang, Y., and Shi, L. (2018). Research on Short-Term Wind Speed Prediction Based on LSTM[J]. *Computer Simulation* 35 (11), 456–461. doi:10.3969/j.issn.1006-9348.2018.11.099

Li, C., Tang, G., Xue, X., Saeed, A., and Hu, X. (2020). Short-Term Wind Speed Interval Prediction Based on Ensemble GRU Model. *IEEE Trans. Sustain. Energ.* 11, 1370–1380. doi:10.1109/TSTE.2019.2926147

Lin, Y., and Liu, P. (2011). Short-Term Wind Power Portfolio Prediction Model Based on Empirical Modal Decomposition and Support Vector Machine[J]. *Chin. J. Electr. Eng.* 31 (31), 102–108. doi:10.13334/j.0258-8013.pcsee.2011.31.014

Liu, A-G., Xue, Y-T., Hu, J-L., and Liu, L-P. (2015). Ultra-Short-Term Prediction of Wind Power Based on GA Optimized SVM[J]. *Power Syst. Prot. Control.* 43 (02), 90–95. doi:10.7667/j.issn.1674-3415.2015.02.014

Liu, H., Mi, X., and Li, Y. (2018). Smart Multi-Step Deep Learning Model for Wind Speed Forecasting Based on Variational Mode Decomposition, Singular Spectrum Analysis, LSTM Network and ELM. *Energ. Convers. Manage.* 159, 54–64. doi:10.1016/j.enconman.2018.01.010

Liu, S., Zhu, Y., Zhang, K., and Gao, J. (2020). Short-term Wind Power Forecasting Based on Error-Corrected ARMA-GARCH Model[J]. *J. Solar Energ.* 41 (10), 268–275. doi:10.1016/j.apenergy.2019.114137

Liu, Y., and Liu, L. (2021). Wind Power Prediction Based on LSTM-CNN Optimization [J]. *Scientific J. Intell. Syst. Res.* 3 (4), 277–285.

Liu, Z., Jiang, P., Zhang, L., and Niu, X. (2020). A Combined Forecasting Model for Time Series: Application to Short-Term Wind Speed Forecasting[J]. *Appl. Energ.* 259, 114137. doi:10.1016/j.apenergy.2019.114137

Louka, P., Galanis, G., Siebert, N., Kariniotakis, G., Katsafados, P., Pytharoulis, I., et al. (2008). Improvements in Wind Speed Forecasts for Wind Power Prediction Purposes Using Kalman Filtering. *J. Wind Eng. Ind. Aerodynamics* 96 (12), 2348–2362. doi:10.1016/j.jweia.2008.03.013

Tascikaraoglu, A., and Uzunoglu, M. (2014). A Review of Combined Approaches for Prediction of Short-Term Wind Speed and Power. *Renew. Sustain. Energ. Rev.* 34, 243–254. doi:10.1016/j.rser.2014.03.033

Wang, L., Dong, L., and Gao, D. (2015). Short-Term Prediction of Wind Power Based on Multi-Position NWP with Principal Component Analysis[J]. *J. Electr. Eng. Technol.* 30 (05), 79–84. doi:10.19595/j.cnki.1000-6753.tces.2015.05.010

Wang, R., Li, C., Fu, W., and Tang, G. (2020). Deep Learning Method Based on Gated Recurrent Unit and Variational Mode Decomposition for Short-Term Wind Power Interval Prediction. *IEEE Trans. Neural Netw. Learn. Syst.* 31 (10), 3814–3827. doi:10.1109/tnnls.2019.2946414

Wang, S., Jiang, X., Zeng, L., and Chang, Y. (2020). Ultra-Short-Term PV Power Prediction Based on VMD-DESN-MSGP Model[J]. *Power Grid Technol.* 44 (03), 917–926. doi:10.1109/TNNLS.2019.2946414

Wang, Y., Wang, J., and Wei, X. (2015). A Hybrid Wind Speed Forecasting Model Based on Phase Space Reconstruction Theory and Markov Model: A Case Study of Wind Farms in Northwest China. *Energy* 91, 556–572. doi:10.1016/j.energy.2015.08.039

Wu, Q., Lu, Y., Wang, Z., and Zhou, Q. (2017). Probabilistic Wind Power Prediction Methods and Prospects[J]. *Power Syst. Automation* 41 (18), 167–175. doi:10.7500/AEPS20160914002

Wu, Z., Xia, X., Xiao, L., and Liu, Y. (2020). Combined Model with Secondary Decomposition-Model Selection and Sample Selection for Multi-step Wind Power Forecasting[J]. *Appl. Energ.* 261 (4), 114345. doi:10.1016/j.apenergy.2019.114345

Wu, Z., Zhao, X., Ma, Y., and Zhao, X. (2019). A Hybrid Model Based on Modified Multi-Objective Cuckoo Search Algorithm for Short-Term Load Forecasting. *Appl. Energ.* 237, 896–909. doi:10.1016/j.apenergy.2019.01.046

Ye, L., and Zhao, Y. (2014). A Review of Research on Wind Power Prediction Based on Spatial Correlation[J]. *Power Syst. Automation* 38 (14), 126–135. doi:10.7500/AEPS20130911004

Zhang, L., Li, J., Wu, Y., Shaojie, B., Kolter, J. Z., Koltun, V., et al. (2016). ARIMA Model Forecast for Wind Power Time Series with Different Temporal Resolutions [J]. *Electric Power* 49 (6), 176–180. doi:10.11930/j.issn.1004-9649.2016.06.176.05

Zhou, W., Zhong, J., Sun, H., Li, G., Kong, J., and Zhang, F. (2018). *An Interval Rolling Estimation Method for Daily Wind Power Forecast Errors Based on Hidden Markov Model*, 90–95.

Conflict of Interest: Authors RS and SW were employed by company State Grid Jiangsu Electric Power Co. Ltd. Research Institute.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The authors declare that this study received funding from State Grid Jiangsu Electric Power Co. Ltd. Research Institute. The funder had the following involvement in the study: interpretation of data.

frontiers in Energy Research

ORIGINAL RESEARCH
published: 17 September 2021
doi: 10.3389/fenrg.2021.720374

Distributed Economic Optimal Scheduling Scheme for Ship-Integrated Energy System Based on Load Prediction Algorithm

Yuxin Zhang[1,2], *Qihe Shan*[1,2]*, *Fei Teng*[3] *and Tieshan Li*[4]

[1]Navigation College, Dalian Maritime University, Dalian, China, [2]Maritime Big Data and Artificial Intelligent Application Centre, Dalian Maritime University, Dalian, China, [3]Marine Electrical Engineering College, Dalian Maritime University, Dalian, China, [4]School of Automation Engineering, University of Electronic Science and Technology of China, Chengdu, China

In order to enhance navigation safety and promote environmental protection, this paper takes the problem of energy management in a ship-integrated energy system into consideration. According to the characteristics of navigation, an intelligent ship energy management model, simultaneously considering the social and economic benefits, has been proposed. Meanwhile, this paper analyzes a distributed optimal scheduling problem which considers renewable generation devices and an energy storage system. Combined with an electricity-power system and thermal-power system, we propose an optimal scheduling scheme to accurately meet the actual load demand based on the pre-results analyzed by the ensemble learning short-term load forecasting algorithm. In addition, the related stability analysis is given. Further, a series of simulation results have been presented, which denote that the proposed load forecasting algorithm can accurately analyze the short-term load demand trend, and the proposed optimization algorithm can effectively coordinate economic and environmental protection.

Keywords: ship integrated energy system, energy management, renewable generation devices, load forecasting algorithm, distributed optimal scheduling, ensemble learning algorithm

OPEN ACCESS

Edited by:
Qiuye Sun,
Northeastern University, China

Reviewed by:
Narottam Das,
Central Queensland University, Australia
Fanlin Meng,
University of Essex, United Kingdom
Geyang Xiao,
Zhejiang Lab, China
Xiting Peng,
Shenyang University of Technology, China

***Correspondence:**
Qihe Shan
shanqihe@dlmu.edu.cn

Specialty section:
This article was submitted to Smart Grids, a section of the journal Frontiers in Energy Research

Received: *04 June 2021*
Accepted: *29 July 2021*
Published: *17 September 2021*

Citation:
Zhang Y, Shan Q, Teng F and Li T (2021) Distributed Economic Optimal Scheduling Scheme for Ship-Integrated Energy System Based on Load Prediction Algorithm. Front. Energy Res. 9:720374. doi: 10.3389/fenrg.2021.720374

1 INTRODUCTION

As we know, a traditional generator which relies on fossil resources will be accompanied by a large amount of greenhouse gases (Sun et al., 2019a; Li et al., 2020a). As one of the areas with the highest consumption of fossil fuels, the shipping industry emits 3–5% of carbon dioxide into the earth every year (Rafiei et al., 2020). In order to reduce the air pollution caused by the shipping industry, the International Maritime Organization (IMO) has issued a series of strict regulations (Czermański et al., 2020), such as limiting the energy efficiency operation index (EEOI) (Fang et al., 2019) and improving energy efficiency, etc. As a new type of ship energy architecture, the ship-integrated energy system (S-IES) improves the utilization efficiency of renewable energy and adjusts the utilization rate of energy, which embodies the deep integration of information technology and energy characteristics. Therefore, with the continuous maturity of intelligent technology (Peng and Wang, 2018; Liang et al., 2020; Li, J. et al., 2020), how to quickly and accurately optimize the scheduling of the ship-integrated energy system, which is based on the concept of sustainable development, has become a hot research topic.

With the continuous development of intelligent technology, its advanced concept has been widely used in various fields (Ye et al., 2018; Liu et al., 2020; Wang et al., 2020; Zhang G. et al., 2020; Lei et al., 2021).

Compared with the traditional ship energy system, to reduce fossil energy consumption and pollution emissions, S-IES combines photovoltaic units (PVs), wind turbine units (WTs) with diesel generators (DGs), and combined heat and power generations (CHPs) to form an energy supply system. In addition, unlike traditional ships which rely on mechanical transmission to provide power, ships based on S-IES are all electric ships (AES) which rely on electric propulsion, further improving the controllability and flexibility of ships. Nowadays, in order to improve the rationality and reliability of the ship-integrated energy system, it is essential to establish an optimal scheduling mechanism model, which can accurately describe the whole voyage state. A hybrid energy optimization management model with wind turbines and energy storage equipment is proposed in Li et al. (2020b), considering charging/discharging efficiency and EEOI throughout the whole voyage, which takes the economic benefit as the primary optimization objective to reduce the operation cost. In order to improve the efficiency of energy utilization and ensure the safe operation of power supply equipment, Kanellos et al. (2016) proposes an energy optimization management model to guarantee that the EEOI of ships in different sailing conditions can meet the requirements of pollution emission limits proposed by IMO. Based on the dual consideration of environmental benefits and economic benefits, according to the actual situation of ship navigation, Wen et al. (2021) proposes a joint optimal scheduling model which considers the shore-side electricity system. At present, the research of ship energy optimal management is mostly based on the constraint of supply and demand balance, and is then used to establish a dynamic relationship between the output power of generators and the actual load demand. Thus, an efficient and accurate load forecasting method is the key to improve the reliability of ship energy optimization management.

According to fuel consumption information, meteorological data, and hydrological information, etc., (Teng et al., 2020), based on a broad learning system, proposes a load forecasting method, which can reduce the negative impact of uncertainty in a complex marine environment during navigation operation. Owing to the new energy generators such as WTs and PVs, increasing the uncertainty of the power system, a prediction intervals (PIs) method based on a neural network is proposed to improve the accuracy of load forecasting (Quan et al., 2014). However, the short-term load of a ship has strong nonlinear variation characteristics, i.e., the ship operation in different periods such as departure, arrival, and cruise has quite different load demands. Thus, the traditional method cannot accurately predict the short-term load variation trend. In addition, in view of the actual navigation operation, it is unable to obtain the above information accurately and in a timely manner. Therefore, based on the actual situation, there are still many challenges in ship short-term load forecasting.

Owing to the above situations, we can transform the optimal scheduling problem of the ship-integrated energy system into an energy planning problem which contains a series of complex navigation safety constraints. Considering the total lifecycle cost of a hybrid electric propulsion ship and the storage performance of the battery, an energy management model which can effectively extend the service life of the ship is proposed in Chen et al. (2020) which is based on the depth of the discharge (DOD) constraint of the battery. Taking the minimum fuel consumption as the optimization objective, a power-flow-based energy management model considering a battery energy storage system (BESS) is established in Balsamo et al. (2020), which can improve energy efficiency and reduce pollution emissions, completing ship energy optimal scheduling simultaneously. Fang et al. (2020) presents an optimal management model for all electric ships (AES), which can consider both environmental and economic benefits. It can ensure the navigation operation of the ship and reduce the investment of additional equipment, e.g., ESS by increasing the ship's power system constraints. However, with the continuous maturity of renewable-related technologies (Farrok et al., 2018), more and more renewable energy supply equipment such as PV units and WT units are connected to the S-IES, which increase the penetration rate of green energy access gradually. And there are a lot of "plug and play" load equipment, such as life load, mechanical, etc. Owing to the above situations, the existing optimization scheduling methods such as the dynamic programming algorithm cannot meet the actual needs of S-IES with strong distributed characteristics. Therefore, it is essential for us to improve the computing speed of the distributed algorithm while dealing with the performance of a ship energy system.

Above all, based on the ship's navigation characteristics, such as carrying capacity, voyage, and velocity, combined with the constraints of the electricity/thermal energy system, such as supply and demand balance, ramping rate constraints, and energy energy-off constraints, this paper proposes a distributed optimal scheduling method for the ship-integrated energy system with load forecasting. The major contributions of this paper are as follows.

1) This paper proposes a distributed ship-integrated energy system with renewable generation devices and an energy storage system, which can obtain environmental and economic benefits at the same time. Combined with the characteristics of ship navigation, such as safety-sailing, greenhouse emission, and the requirements of electricity/thermal load during the whole voyage, we take the lowest voyage operation cost as the primary optimization objective to form a distributed energy optimization management model for multi-energy ships, so as to realize the economic optimization scheduling of the whole voyage.
2) In order to reduce the waste of fossil energy caused by redundant capacity and improve the accuracy of load response in a ship-integrated energy system, this paper proposes an intelligent algorithm which can quickly and accurately predict the load of navigation. In addition, for improving the efficient and reasonable utilization of energy, a distributed optimal scheduling algorithm for the ship-integrated energy system is proposed as well, which can guarantee safe and reliable navigation and quickly respond to the actual demand of the load equipment simultaneously.

The remainder of this paper is organized as follows. In **Section 2**, we identify the main features and structures in the ship-integrated energy system. **Section 3** proposes an energy

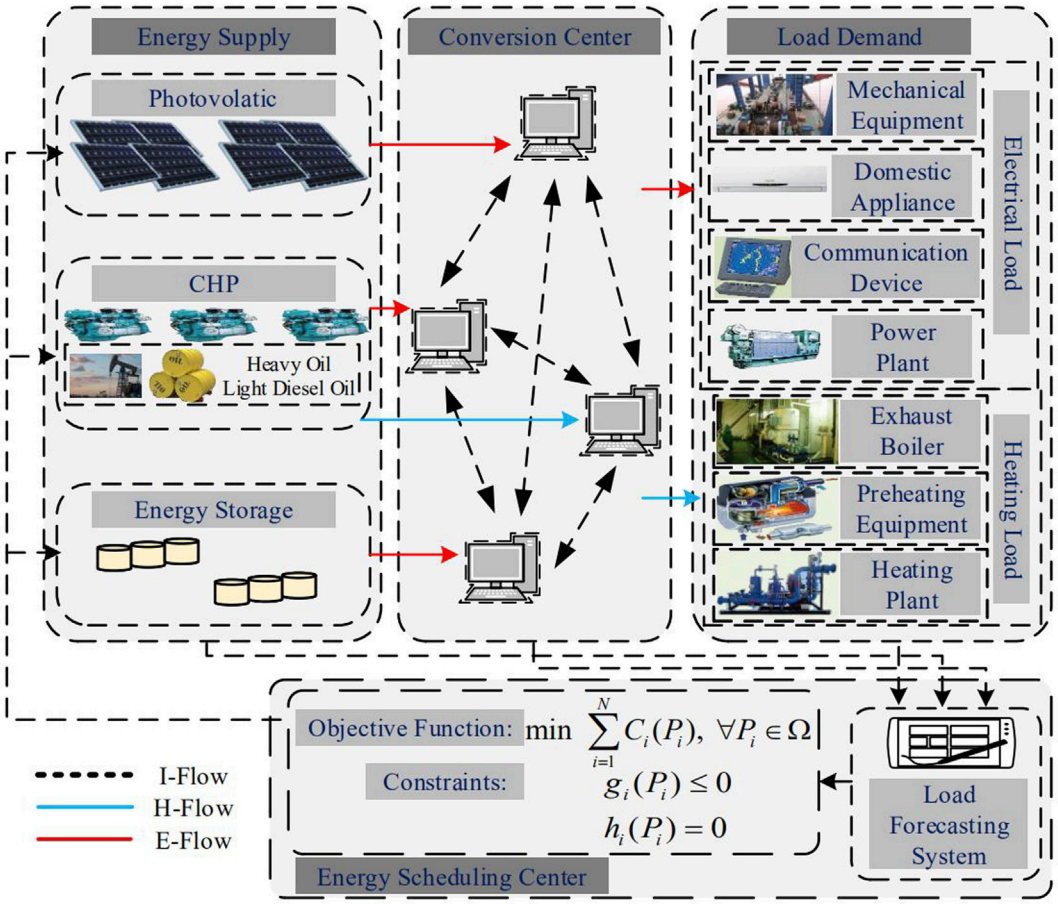

FIGURE 1 | Basic structure of the integrated energy system on intelligent ships.

management scheme based on load forecasting. **Section 4** presents a distributed economic optimal scheduling algorithm for S-IES. According to the analysis of the example and simulation results, the correctness of the proposed viewpoint is fully proved in **Section 5**. **Section 6** summarizes and concludes the points from the paper.

2 STRUCTURE AND FEATURES IN S-IES

The intelligent ship-integrated energy system considered in this paper can accurately predict short-term load demand according to historical navigation information, and owing to the prediction data, combined with the working characteristics of the ship, we can optimize the energy output of the energy supply equipment. **Figure 1** shows the basic architecture of the S-IES. According to **Figure 1**, the S-IES can be roughly divided into an energy supply system, load demand system, energy conversion center, load forecasting center, and energy optimal management center. An energy supply system provides electricity and thermal power for the ship load to maintain normal operation during different conditions; as an energy router, the energy conversion center processes information flow and energy flow simultaneously, and undertakes the task of mutual conversion of electricity and thermal energy; based on navigation constraints and energy constraints, the energy management center adopts a distributed intelligent algorithm to analyze and calculate the energy optimal scheduling scheme in a short period.

2.1 Velocity and Voyage Modeling

Ship navigation can be divided into three different operation conditions, i.e., cruising, anchoring, and docking. **Figure 2** is a typical cross section of the ship, in which the voyage and velocity range limits of the ship are indicated. As we know, in the course of sailing, we can adjust the velocity within a certain range based on the actual demand, but the ship must arrive at the ports or intermediate-ports at the specified time. In addition, generally, the ship speed depends on the propulsion power which has a certain relationship with the resistance of sailing. If the propulsion power has a period of t, the velocity V^t can be described as

$$V^t = \sqrt[pr_2]{\frac{L^t_{pr}}{pr_1}} \tag{1}$$

where, pr_1 is a nonnegative constant, which is related to the hull form type, generally taken as 3 (Kanellos et al., 2014); pr_2 is the matching parameter of propulsion power and ship speed. Owing

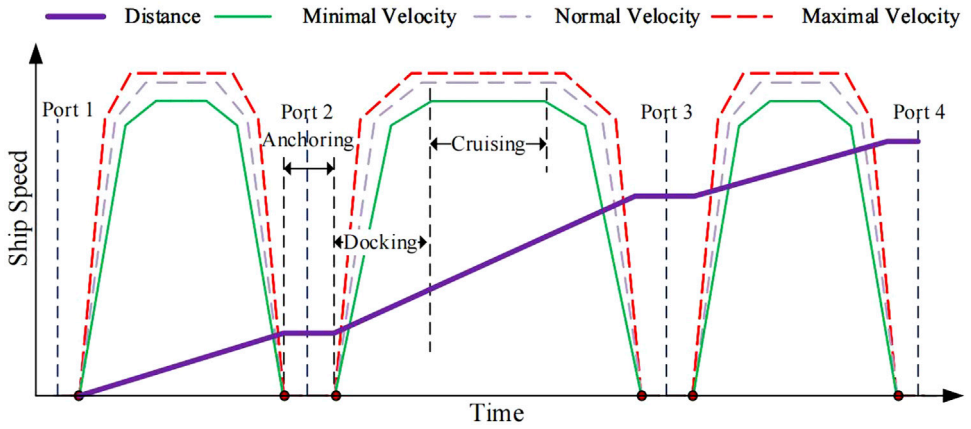

FIGURE 2 | Typical cross section of a sailing voyage with velocity limitation.

to above situations, the mathematical expression of voyage and velocity range limits are as follows, which is given in Feng et al. (2018).

$$(1-\sigma)\cdot V_n^t \le V^t \le (1+\sigma)\cdot V_n^t \tag{2}$$

$$Dist^t = \begin{cases} Dist^{t-1} + V^t \cdot \Delta t, & t > 1 \\ V^t \cdot \Delta t, & t = 1 \end{cases} \tag{3}$$

where, V_nt denotes the ship speed at t; σ represents the interval coefficient of ship speed; $Dist^{t-1}$, $Dist^t$ represent the voyage at a period of $t-1$ and t, respectively; Δt represents the specified time interval.

2.2 Energy Supply System Modeling

As mentioned before, in order to improve the environmental friendliness of the ship, we integrate PVs and CHPs into the S-IES. In addition, owing to the high-power ramps and the intermittency caused by renewable energy equipment, a battery energy storage system (BESS) has been considered in the ship to ensure the stability of power output.

1) CHP System Modeling: With the improvement of energy saving and efficiency awareness, CHP, as the preferred equipment which can reduce consumption and increase efficiency, has been widely applied. CHP can utilize the after-heat generated by power generation to other heating loads, such as fuel preheating, etc. It has been considered as the most energy-saving method for power generation, thus CHP has a tendency to be installed on intelligent ships. The operation cost function of CHP, containing fuel consumption, has been modeled as the following convex function in Sun et al. (2019b).

$$\begin{aligned} C\left(P_t^{CHP}, H_t^{CHP}\right) = \sum_{i=1}^{n} a_{i,1}\cdot\left(P_{i,t}^{CHP}\right)^2 + a_{i,2}\cdot P_{i,t}^{CHP} + \\ b_{i,1}\cdot\left(H_{i,t}^{CHP}\right)^2 + b_{i,2}\cdot H_{i,t}^{CHP} + c_i\cdot\left(P_{i,t}^{CHP}\cdot H_{i,t}^{CHP}\right) + \kappa \end{aligned} \tag{4}$$

where, n is the number of CHPs; $P_{i,t}^{CHP}$ denotes the electrical output of the ith CHP at the period of t; $H_{i,t}^{CHP}$ represents the heating output of theith CHP at the period of t; $a_{i,1}$, $a_{i,2}$, $b_{i,1}$, $b_{i,2}$, c_i, and κ are the operating cost parameters of the ith CHP.

2) PV Modeling: In order to improve the efficiency of energy utilization and reduce the pollution emissions during the sailing voyage, PV units have been integrated in the S-IES. In addition, the production capacity of a PV panel is not only related to the strength of illumination, degree of angle and panel area, but also related to the ship route and deck inclination angle because of the ship characteristics. Owing to these situations, Long et al. (2020) and Wen et al. (2020) present a mathematical model which can describe the capacity of PV units on ships appropriately.

$$\begin{aligned} P_t^{PV} &= \sum_{i=1}^{m} \eta_i^{PV}\cdot Area_i^{PV}\cdot I_t^{PV}\cdot \varepsilon_{i,t}^{PV} \\ \eta_i^{PV} &= \eta_{ref}^{PV}\cdot \eta_{MPPT}\left[1-\beta\left(T_{i,PV}-T_{PV-ref}\right)\right] \\ \varepsilon_{i,t}^{PV} &= cos\theta_{i,t} + \mu_{i,1}cos\left(\phi_t/2\right)^2 + \mu_{i,2}sin\left(\phi_t/2\right)^2 \end{aligned} \tag{5}$$

where, P^{PV} denotes the power output of PV; η^{PV}, η_{ref}^{PV} represent the efficiency and reference efficiency of photovoltaic panels, respectively; $Area^{PV}$ denotes the PV panel area; the radiation intensity can be described as I^{PV}; η_{MPPT} represents the tracking efficiency; T_{PV}, T_{PV-ref} denote the temperature and reference temperature of PV panels, respectively; the angle between PV panels and radiation can be used as θ; ϕ is the tilt angle; μ_1 and μ_2 denote the angle parameters.

3) BESS Modeling: Considering that there are many factors affecting solar energy, taking PV units into the energy system directly will lead to S-IES volatility. In addition, the ship energy system will be unstable influenced by the operation of high-load equipment and sudden speed-adjusting. Owing to the above situations, we integrate BESS into the S-IES as additional energy supply generation which can share the pressure with the other device (Gangatharan et al., 2020). The model of charging and discharging can be expressed as follows (Zhang, Y. et al., 2020).

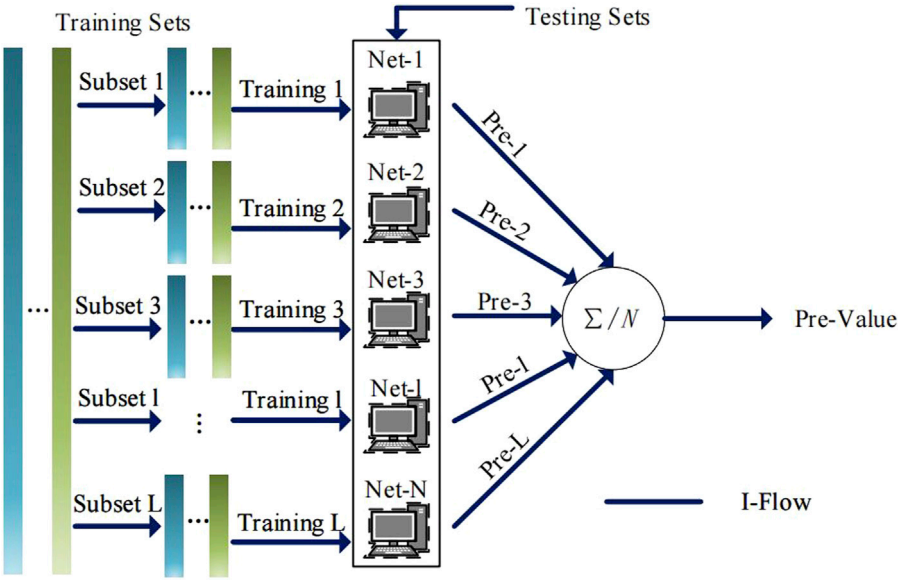

FIGURE 3 | Structure of load forecasting algorithm.

$$E_t^{SS} = \begin{cases} E_{t-1}^{SS} + P_t^{ESS} \cdot \Delta t, & t > 1 \\ E_{initial} + P_t^{ESS} \cdot \Delta t, & t = 1 \end{cases}$$
$$SoC_t^{ESS} = E_t^{SS}/Cap \tag{6}$$
$$SoC_{t-1}^{ESS} + P_t^{ESS} \cdot \eta_t^{ch} \le SoC_{max}^{ESS}$$
$$SoC_{t-1}^{ESS} - P_t^{ESS}/\eta_t^{dis} \ge SoC_{min}^{ESS}$$

where, P_t^{ESS} is the power outputs of BESS; E_t^{SS}, E_{t-1}^{SS} denote the capacity of BESS during the period of t and $t-1$, respectively; SoC_t^{ESS}, SoC_{t-1}^{ESS} represent the state of charge during the period of t and $t-1$; Cap is the fixed capacity; η_t^{dis}, η_t^{ch} are the efficiency of discharging and charging, respectively.

3 OPTIMAL ENERGY DISPATCH SCHEME AND LOAD FORECASTING ON THE S-IES

Energy supply side of the S-IES proposed in this paper consists of PVs, CHPs, and BESS. According to the value of load prediction, the energy management center can obtain an optimal scheduling scheme within sailing constraints. Assume that the S-IES contains n-CHPs, m-PVs, a BESS, and other load equipment, such as service load, propulsion load, mechanical load, etc.

3.1 Load Forecasting Algorithm

In order to realize the optimal scheduling of the S-IES, we need to know the short-term load-demand information in a timely manner. Therefore, we need to use the relevance of specific routes to make a predictive analysis of the load-demand information, and make a reliable prediction of the trend of ship load demand. A deep learning algorithm can collect feature information from ship historical load data through multiple hidden layers, but it is easy to fall into a local optimal solution and other problems to some extent (Zhang, J. et al., 2020; Wang et al., 2021). In order to solve the above contradictory problems, the ensemble learning theory, containing the Bagging method and Bossting method which are two popular algorithms based on resampling for load prediction problems Feng et al. (2018) has received widespread consideration because of its excellent performance in numerical prediction situations. Therefore, this paper proposes a ship short-term load forecasting algorithm based on ensemble learning; the basic structure and framework can be seen in **Figure 3**.

Considering the proposed load prediction algorithm, the historical load data set can be arranged as X^T, firstly. And then, we can obtain the load data subsets X_l by resampling which needs to ensure the samples are not related (i.e., the factors within it without any repeated data), where $l = 1, 2, \ldots, L$. It is worth noting that in the process of sampling and resampling for historical ship load, the time dependence of load data sets should be guaranteed at all times. In addition, for increasing the accuracy and reliability of the proposed algorithm, this paper utilizes Moving Block Bootstrap (MBB) during the process of resampling. Suppose load data subset X^l has H load data factors and samples $H - l_X + 1$ times continuously and repeatedly, where l_X denotes the length of resampling load data subsets. Then, the sampling process of ship historical load data at this time can be seen in **Figure 4** and **Table 1**, the resampling load data subsets can be represented as follows.

$$X^l = \left(x_1^l, x_2^l, \ldots, x_{H-l_X+1}^l\right), l = 1, 2, \ldots, L \tag{7}$$

According to the ship-load forecasting problem, we need to repeat the above processes L times to obtain the independent load data training subset $X = [X^1, X^2, \ldots, X^L]$. $f_l^1(\cdot)$, $f_l^2(\cdot)$, $f_l^{H-l_X+1}(\cdot)$ are the corresponding predicted models which can be regarded as weak predictors as well. Owing to the above models, we can calculate a set of ship-load forecasting values. In addition, there are $L(H - l_X + 1)$

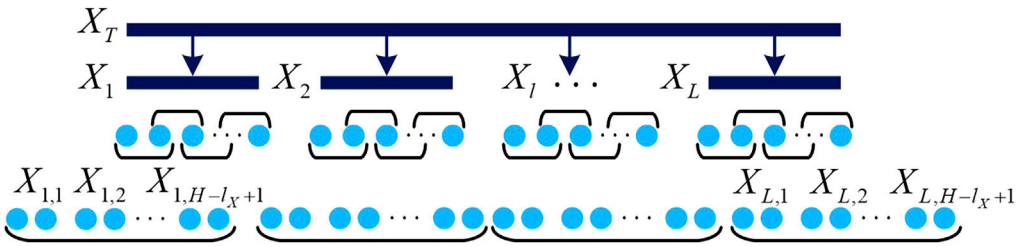

FIGURE 4 | Structure of resampling ship-load subsets with MBB.

predicted models, and we can obtain the corresponding forecasting ship-load values by them. Thus, according to the proposed ship short-term load forecasting algorithm, the final predicted ship-load value can be analyzed by a constructed stronger prediction model which can add them up and take an average. Mathematical expression about the final forecasting short-term ship load $Load_l^{Pre}$ can be represented as follows.

$$Load_l^{Pre} = \frac{\sum_{l=1}^{L} f_l^1(x_l^1) + f_l^2(x_l^2) + \cdots + f_l^{H-l_X+1}(x_l^{H-l_X+1})}{L(H-l_X+1)} \quad (8)$$

3.2 Optimal Energy Dispatch Scheduling Scheme

3.2.1 Optimal Goal Function

The S-IES proposed in this paper considers the environmental and economic benefits comprehensively. And proposes an optimal management mechanism which takes the lowest operation cost as the primary optimization objective, to meet the various load-demand constraints during different working conditions such as berthing and cruising. Owing to the generation equipment installed in the ship, in this paper, the objective function can be separated into three parts which contain the CHPs-cost, PVs-cost, and excitation parameters of BESS. The mathematical model can be represented as the following form.

$$\begin{aligned} &min\{C_{CHP} + C_{PV} + \Upsilon_{BESS}\} \\ &C^{CHP} = \sum_{t=1}^{N_T} C(P_t^{CHP}, H_t^{CHP}) \\ &C^{PV} = \sum_{t=1}^{N_T} \sum_{j=1}^{m} d_{j,1} \cdot \left(P_{j,t}^{PV}\right)^2 + d_{j,0} \\ &\Upsilon_{BESS} = c_0 \cdot P_t^{BESS} \end{aligned} \quad (9)$$

where, C^{CHP}, C^{PV} are the operation cost of CHPs and PVs; N_T denotes the sum of time slots; $d_{j,1}$, $d_{j,0}$ are the operating parameters of PVs;c_0 denotes the influence factor; Υ_{BESS} can be expressed as a penalty coefficient which can enhance the participation of BESS and improve the energy efficiency.

Remark. Since the above mentioned equipment has a longer service life, we assume that they have little loss during the energy optimization management scheduling in the single sailing voyage. Therefore, the cost of equipment maintenance does not need to be considered in the optimization objective function.

TABLE 1 | Ship load forecasting algorithm.

Process of ship electricity/thermal-load prediction
Step 1: Summarize the historical data information of ship electricity/thermal load as X^T
Step 2: Intercept the historical load data set X^T with Lth, randomly. $X^T = \{X_1, X_{i,2}, \ldots, X_L\}$
Step 3: Resample X_l, $l = 1, 2, \ldots, L$ with MBB.
Step 4: Establish weak predictors
Step 5: Obtain corresponding pre-load by weak predictors with X_l for $L(H - l_X + 1)$ times
Step 6: Calculate the final forecasting ship load
Step 7: End

3.2.2 Constraints

Considering the particularity of ship sailing voyages, this paper does not only consider the conventional constraints in energy management such as the power balance constraint and power output constraint, but also the navigation constraints **Eqs 1-3**, interruption of electricity/heating constraint, ramping-rate constraint, and energy efficiency operation index (EEOI) constraints. The specific descriptions are as follows.

1) Electricity/Heating Power Balance Constraints: In order to ensure the normal operation of all load equipment such as the propulsion system and service load, we need to guarantee that the sum of the electrical power output of CHPs, PVs, and ESS matches the electrical load, and the thermal power output of CHPs can meet the thermal-load demand. In addition, the energy conversion center undertakes the mutual conversion of electricity and heating power, i.e., when the thermal output cannot meet the load demand, the electric energy could be converted into thermal power as an additional heating supplier to ensure the normal operation of heating-load equipment. We can express the balance constraints as the following.

$$\begin{aligned} &\sum_{i=1}^{n} P_{i,t}^{CHP} + \sum_{j=1}^{m} P_{j,t}^{PV} + P_t^{ESS} = L_{P,t} + \Delta L_{P,t} \\ &\sum_{i=1}^{n} H_{i,t}^{CHP} + \Delta L_{H,t} = L_{H,t}, \Delta L_{H,t} = \eta_{EX} \Delta L_{P,t} \end{aligned} \quad (10)$$

where, $L_{P,t}$, $L_{H,t}$ represent the electricity and thermal load value at the period of t, respectively; η_{EX} is the conversion efficiency parameter; $\Delta L_{P,t}$ denotes the conversion of electricity into

TABLE 2 | Operation cost parameters of 3-CHPs and 4-PVs.

	$a_{i,1}$, \$/MW^2	$a_{i,2}$, \$/MW	$b_{i,1}$, \$/MW^2	$b_{i,2}$, \$/MW	c_i, \$/MW^2	κ, \$
CHP-1	250	2000	200	1500	150	20
CHP-2	200	1900	210	1800	200	20
CHP-3	200	1900	210	1800	200	20
	$d_{i,1}$, \$/MW^2	$d_{i,0}$, \$			$d_{i,1}$, \$/MW^2	$d_{i,0}$,\$
PV-1	300	29		PV-3	210	10
PV-2	291	19		PV-4	210	10

thermal load value; $\Delta L_{H,t}$ is the thermal power supplied by electricity generators.

Remark. When BESS is discharging as a power supply equipment, P_t^{ESS} has a positive value; on the contrary, when BESS is charging as a load equipment, P_t^{ESS} has a negative value. In addition, the value of short-term load-demand prediction can be obtained by the load forecasting algorithm proposed in this paper through the analysis and calculation of historical sailing data information.

2) Power Outputs Constraints: It is noted that we should ensure the reliable and safe working conditions of the power-supply equipment during the whole sailing voyage. Therefore, it is essential to restrict its output power to improve the safety of the energy management model in the S-IES. In this paper, combined with the actual ship situations, the minimum and maximum limits on the power outputs of each equipment are as follows.

$$\begin{aligned} &P_{i,min}^{CHP} \le P_{i,t}^{CHP} \le P_{i,max}^{CHP}, H_{i,min}^{CHP} \le H_{i,t}^{CHP} \le H_{i,max}^{CHP} \\ &P_{j,min}^{PV} \le P_{j,t}^{PV} \le P_{j,max}^{PV}, P_{min}^{BESS} \le P_t^{BESS} \le P_{max}^{BESS} \end{aligned} \tag{11}$$

where, $P_{i,max}^{CHP}$, $P_{j,max}^{PV}$, P_{max}^{BESS} are the maximal electricity power outputs of CHPs, PVs, and BESS, respectively; $P_{i,min}^{CHP}$, $P_{j,min}^{PV}$, P_{min}^{BESS} denote the minimum electricity power outputs working state of CHPs, PVs, and BESS, respectively; $H_{i,max}^{CHP}$ $H_{i,min}^{CHP}$ are the maximal and minimal thermal power outputs value of CHPs. In addition, $P_{i,max}^{CHP}$, $P_{j,max}^{PV}$, P_{max}^{BESS}, $P_{i,min}^{CHP}$, $P_{j,min}^{PV}$, $H_{i,max}^{CHP}$, $H_{i,min}^{CHP}$ are positive constants and P_{min}^{BESS} is a negative constant determined by the operating performance of each equipment.

3) Electricity and Thermal Power-off Constraints: Considering the difference between the traditional power system, the S-IES should ensure that the critical equipment operates normally during the whole voyage owing to the navigation characteristics. Based on the consideration of ship safety, we take the restriction of electricity and thermal power output as an additional constraint, i.e., the anti-power-off constraint. The mathematical model is as follows.

$$\begin{aligned} L_{P,M} &\le \sum_{i=1}^{n} P_{i,max}^{CHP} + \sum_{j=1}^{m} P_{j,max}^{PV} + P_{max}^{BESS} - max\left\{P_{i,max}^{CHP}, P_{j,max}^{PV}, P_{max}^{BESS}\right\} \\ L_{H,M} &\le \sum_{i=1}^{n} H_{i,max}^{CHP} + \Delta L_{H,max} - max\left\{H_{i,max}^{CHP}, \Delta L_{H,max}\right\} \end{aligned} \tag{12}$$

According to **Eq. 10**, we can confirm that the electricity/thermal power can meet the must-run load demand while the highest capacity equipment breaks down suddenly. $L_{P,M}$ is the must-run electricity load such as a propulsion system, communication and navigation equipment; $L_{H,M}$ is the must-run thermal load such as fuel preheating equipment; $\Delta L_{H,max}$ denotes the maximal thermal power generated by the electricity power conversion. Owing to **Eq. 12**, we can ensure that must-run load devices such as communication equipment, propulsion system, can maintain normal operation when the highest capacity generator breaks down.

4) EEOI Constraints: EEOI, as a crucial detection parameter, is essential to evaluate the greenhouse gas emissions during ship navigation. Therefore, in order to improve the environmental protection of sailing operation, the EEOI constraint is added in this paper to ensure that the pollution emission of each navigation period is lower than the preset value. The mathematical equation on the EEOI constraint (Kanellos, 2014) can be expressed as follows.

$$\begin{aligned} EEOI_1 &= \frac{CO_2}{M_{Load} \cdot Dist} \\ &= \frac{\sum_{i=1}^{n}\left(c_{2,i} \cdot \left(P_{i,t}^{CHP}\right)^2 + c_{1,i} \cdot P_{i,t}^{CHP} + c_{0,i}\right)}{M_{Load} \cdot Dist} \end{aligned}$$

where, $EEOI_1$ is the ship index during sailing; CO_2 denotes the value of carbon dioxide emissions; M_{Load} means the ship carrying capacity; $c_{2,i}$, $c_{1,i}$, and $c_{0,i}$ are the coefficients between power outputs and greenhouse gas emission. It is noted that the carbon emission function is a quadratic convex, therefore, this paper enlarges the molecular part to obtain a better form of calculation, i.e.,

$$EEOI_1 \le \frac{\sum_{i=1}^{n}\left(d_{2,i} \cdot P_{i,t}^{CHP} + d_{0,i}\right)}{M_{Load} \cdot Dist} \le EEOI_{Set} \tag{13}$$

4 DISTRIBUTED ECONOMIC OPTIMAL SCHEDULING SCHEME ON THE S-IES

Owing to the energy system considered, this paper installed a large number of new energy equipment, so the system has strong distributed characteristics. In addition, the energy system and communication network have been integrated in the system with the power outputs and load-demand information transmitted among the devices. Based on the above situations, this paper proposes a distributed economic optimal scheduling scheme for

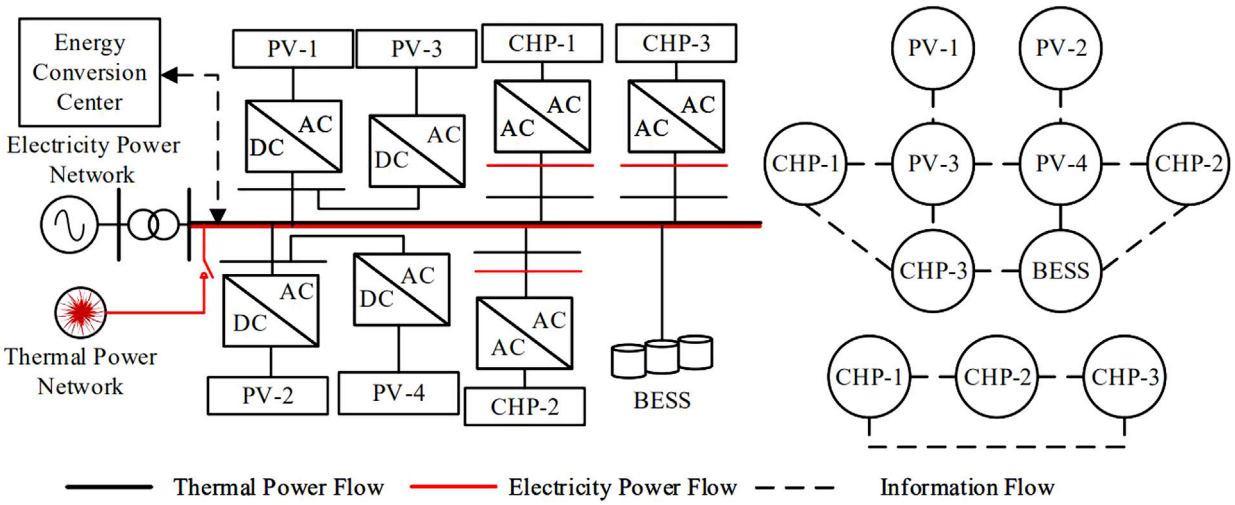

FIGURE 5 | Physical topology and information communication topology in the S-IES.

TABLE 3 | Error comparison between NN and proposed forecasting algorithm based on 200 sets.

	A-MAE (kW)	A-MSE	A-RMSE	Total absolute error (MW)
NN algorithm	11.0	0.0013	0.0276	77.9738
Proposed algorithm	10.8	0.0012	0.0271	77.2005

the S-IES, which ensures the system can meet the load demand and obtain the minimal system operation cost, as well.

4.1 Graph Theory Infrastructure

In this paper, $V = \{1, 2, \ldots, N\}$ is regarded as a set of nodes in the system, and the edge set can be expressed as $E \subseteq V \times V$. Therefore, based on the edge set and node set, the communication topology in the system can be expressed as $\Xi = (V, E)$. If node i and node j are neighbors to each other, then we can describe the neighbor node set of i as $N_{ij} = \{j \subseteq V | (V_i, V_j) \subseteq E\}$ (Ma et al., 2021). And the connection weight between two nodes can be expressed as ω_{ij}, which should ensure that $\omega_{ij} > 0$ when two nodes are neighbors; $\omega_{ij} = 0$ when there is no communication between two nodes. If the communication structure Ξ considered in this paper is undirected and connected, i.e., there is a path between any two nodes then $\omega_{ij} = \omega_{ji}$.

RemarkAccording to the above theory, we can obtain that $\sum_{j=1}^{N} \omega_{ij} = 1$ and $\sum_{i=1}^{N} \omega_{ji} = 1$. It is noted that the graph theory influences the device connection weight during the period of information exchange. When devices are connected to each other, their connection weight is greater than zero. On the contrary, the connection weight is zero when the devices are not connected, which means they cannot exchange their state information (power-outputs, operation error, Lagrange multipliers, etc.) to each other. In this paper, we assume that $\omega_{ij} = \frac{1}{|N_i|+|N_{ij}|+\rho}$ when $i \neq j$ and $\omega_{ij} = 1 - \frac{1}{|N_i|+|N_{ij}|+\rho}$ when $i = j$, which is influenced by the number of itself and its neighboring two parts, where $|N_i|$ denotes the number of energy-suppliers; $|N_{ij}|$ is the neighbors' number of node i; ρ is a minimal positive constant which ensures the denominator cannot be zero.

4.2 Problem Statement and Distributed Optimization Algorithm

The optimal scheduling model on the S-IES mentioned above is based on the balance of supply and demand which establishes the relationship between economic benefits and ship operation constraints. Generally, the optimal scheduling in this paper can be described as a problem on $min\{C_{CHP} + C_{PV} + \Upsilon_{BESS}\}$. The alternating direction method of multipliers algorithm (ADMM) could be used for energy management problems with large-scale constraints, and it is decomposable, as well (Zhang et al., 2017). Therefore, this paper proposes a distributed economic optimal scheduling scheme based on ADMM, which will be introduced from three aspects, i.e., problem transformation, optimizing the iterative process, and convergence judgment.

4.2.1 Problem Transformation

Owing to the fact that the ADMM algorithm cannot deal with the optimal scheduling problem with inequality constraints (Falsone et al., 2021), this paper introduces relaxation variables $S = \{S_1, \ldots, S_m\}$ in the energy management. Thus, inequality constraints can be changed into equality constraints as $HQ = L$, where $Q = \{P_1, \ldots, P_n, S_1, \ldots, S_m\}$.

4.2.2 Optimizing the Iterative Process

Based on the distributed communication structure, the proposed algorithm will optimize the power outputs of power-supply equipment, equipment operation errors, and the Lagrange multipliers. The $K + 1$th iteration expression about the above

TABLE 4 | Parameters of electricity/thermal power outputs and initial value.

	E-min, MW	E-max, MW	T-min, MW	T-max, MW	Initialvalue, MW
CHP-1	6	20	3	15	6/3
CHP-2	6	20	3	15	6/3
CHP-3	6	20	3	10	6/3
PV-1	5	10	—	—	5/-
PV-2	5	10	—	—	5/-
PV-3	5	10	—	—	5/-
PV-4	5	10	—	—	5/-
BESS	−10	10	—	—	0/-

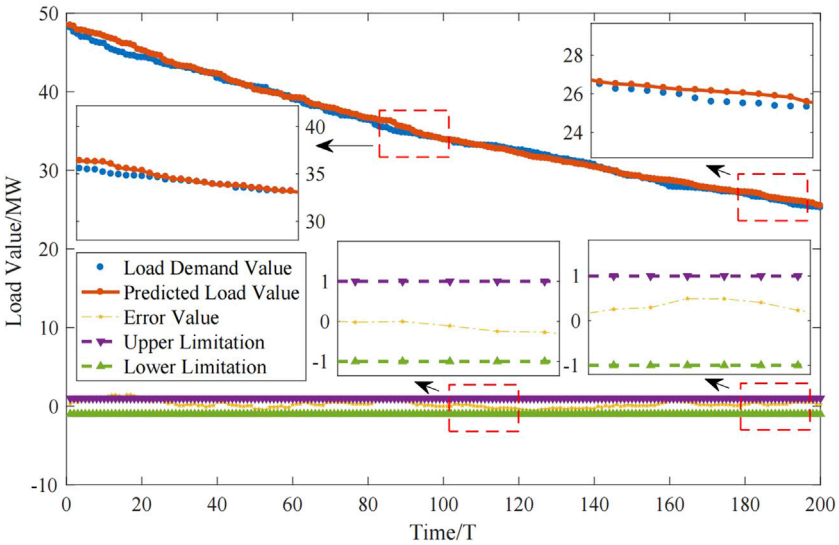

FIGURE 6 | Load forecasting trajectory of electricity power.

variables are as follows, where c denotes the step size; A_d is the parameter of the coupling constraint; ΔQ_K represents the operation error.

$$
\begin{aligned}
&Q_{k+1} \in \arg\min_{Q\in\Omega}\left\{C(Q) + (W\mu_k)^T A_d Q + \frac{c}{2}\|A_d Q - A_d Q_k + W\Delta Q_k\|^2\right\}\\
&\Delta Q_{k+1} = W\Delta Q_k + A_d Q_{k+1} - A_d Q_k\\
&\mu_{k+1} = W\mu_k + c\Delta Q_{k+1}
\end{aligned}
\tag{14}
$$

4.2.3 Convergence Judgment

It is noted that the objective function C considered in this paper is convex and the set Q is convex and compact. In addition, we assume that we can find at least one existing saddle point in the above optimal scheduling problem. According to the above assumptions, we find that the sequences of $\{\Delta Q_k - \Delta\bar{Q}_k\}$, $\{\mu_k - \bar{\mu}_k\}$, $\{Q_k\}$, and $\{A_d Q_k - \Delta\bar{Q}_k\}$ are all bounded. In addition,

$$
\begin{aligned}
&\lim_{k\to\infty}\{\Delta Q_k - \Delta\bar{Q}_k\} = 0\\
&\lim_{k\to\infty}\{\mu_k - \bar{\mu}_k\} = 0\\
&\lim_{k\to\infty}\{\Delta\bar{Q}_k\} = 0
\end{aligned}
$$

Thus, we find that the sequence $\{\|\bar{\mu}_k - \mu^*\|^2 + c^2\|A_d Q_k - A_d Q^* - \Delta\bar{Q}_k\|^2\}_{k\geq 0}$ is convergent, where Q^* and μ^* are the optimal solutions of Q and μ.

5 CASE STUDY

According to the simulation experimental information of the short-term load forecasting algorithm introduced in **Section 3.1**, this chapter will optimize the scheduling of the S-IES based on the proposed distributed optimal economic scheme, and verify the effectiveness according to the simulation experimental results.

5.1 Introduction to the Simulation System

This chapter considers a simulation system which contains 3-CHPs, 4-PVs, and a BESS. According to the short-term load forecasting algorithm and the distributed optimization scheme, generation units (i.e., CHPs, PVs, BESS) will provide energy to the electricity-power network and the thermal-power network to meet the load demand. In addition, each device can exchange power information with their neighbors. The physical and information topology on

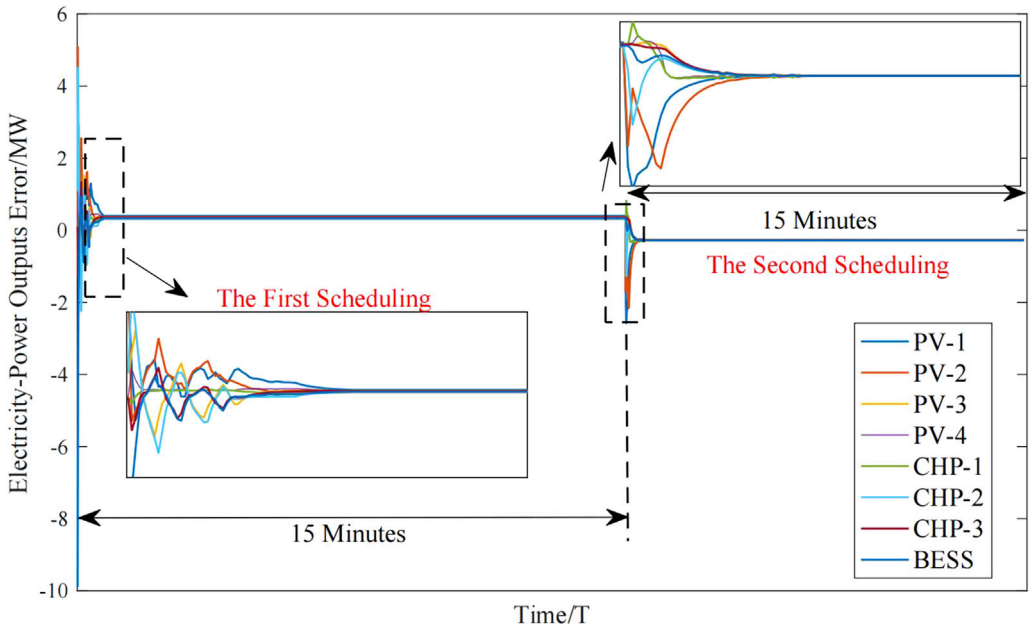

FIGURE 7 | Electricity-power outputs error on generators.

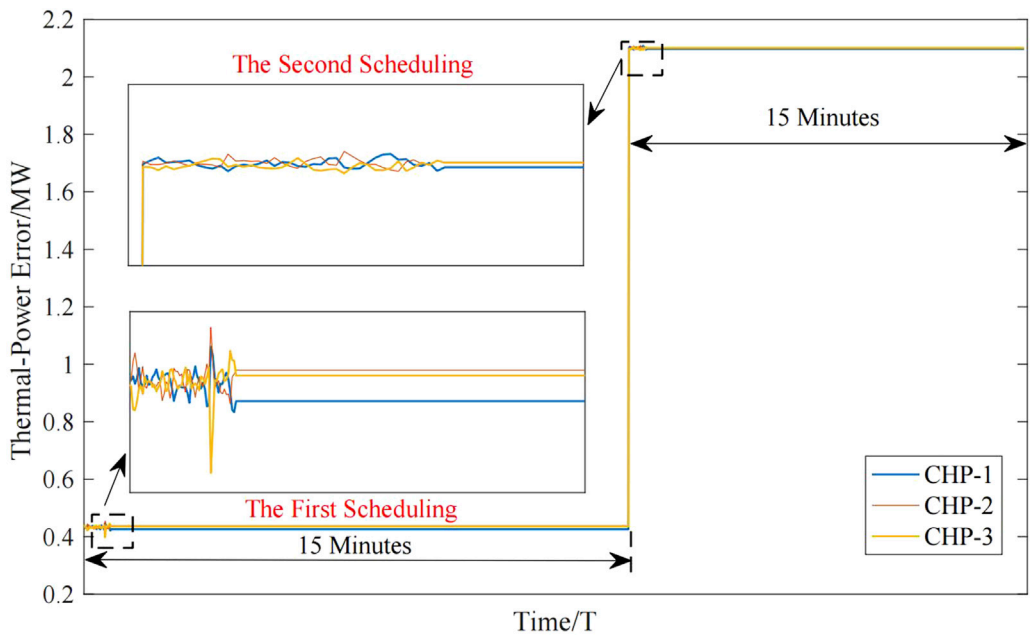

FIGURE 8 | Thermal-power outputs error on generators.

electricity/thermal power can be seen in **Figure 5**. The operation cost parameters are shown in **Table 2**.

5.2 Analysis on Simulation Results

5.2.1 Ship-Load Forecasting Algorithm Analysis

Based on the data samples of historical ship load, the algorithm intercepts the data repeatedly first, which obtains the ship-load resampling subsets. According to the above load data sets, training networks (i.e., weak predictor models) based on BPNN are established in the case study. Meanwhile, the corresponding predicted ship load can be calculated, and then we can obtain the final forecasting ship-load value by the stronger predictor which adds them up and takes an average of them. According to the above process, we can increase the weight of the weak predictor with a smaller load forecasting error and

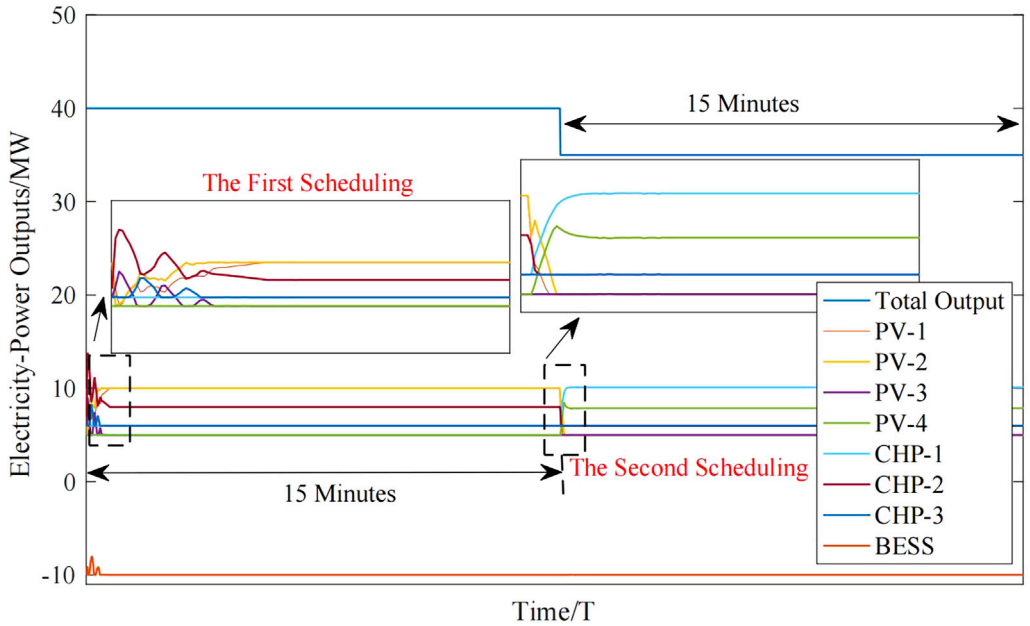

FIGURE 9 | Electricity-power outputs on generators.

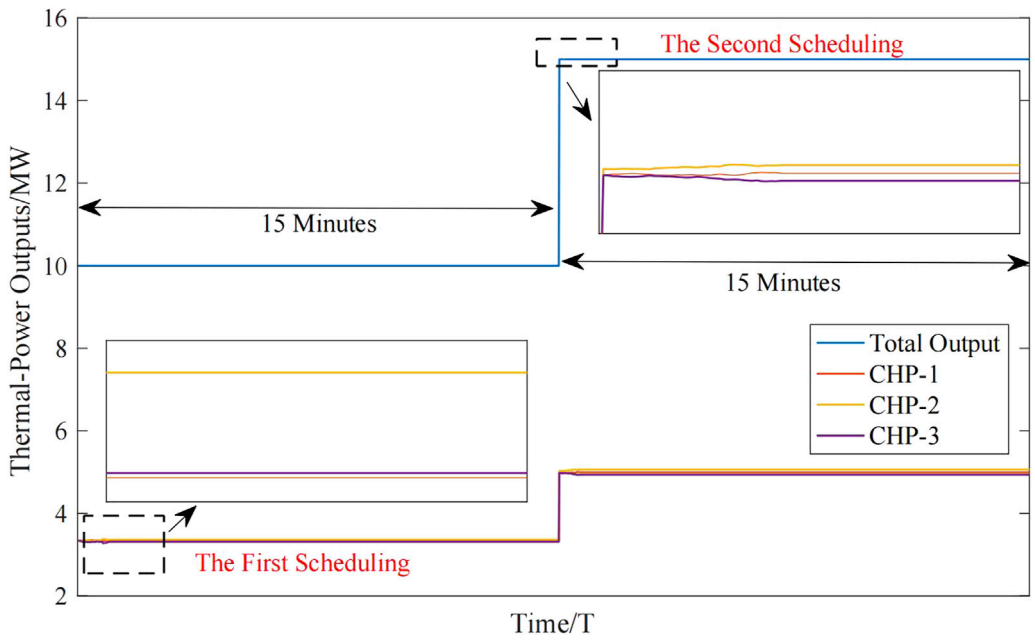

FIGURE 10 | Thermal-power outputs on generators.

reduce the weight of the weak predictor with a large load forecasting error.

This paper selects the historical electricity/thermal load data from the intelligent ship during the whole sailing voyage (Teng et al., 2020) to analyze the short-term load forecasting model. According to the proposed algorithm, the output trajectory of ship-load forecasting results can be obtained by simulation which is shown in **Figure 6**. It can be seen from the figure that the two curves of load forecasting trajectory and real load trajectory have obvious similarity. In addition, this paper takes the NN method for load forecasting as a comparative algorithm. Owing to the simulation, **Figure 6** and the error comparison in **Table 3** can verity that the proposed short-term ship-load forecasting algorithm can accurately predict the trend of ship-load demand for the S-IES.

TABLE 5 | Results summary of the two scheduling schemes.

The first scheduling scheme				The second scheduling scheme			
Results	**E-Power, MW**		**T-Power, MW**	**Results**	**E-Power, MW**		**T-Power, MW**
	Each	**Total**	**Each**		**Each**	**Total**	**Each**
CHP-1	6.0000	19.9999	3.3140	**CHP-1**	10.0760	22.1071	5.0019
CHP-2	7.9999		3.3328	**CHP-2**	6.0104		5.0252
CHP-3	6.0000		3.3532	**CHP-3**	6.0207		4.9729
PV-1	10.0000	30.0001	–	**PV-1**	5.0070	22.8722	–
PV-2	10.0000		–	**PV-2**	5.0069		–
PV-3	5.0001		–	**PV-3**	5.0104		–
PV-4	5.0000		–	**PV-4**	7.8479		–
BESS	–10.0000	Charging	–	**BESS**	–9.9793	Charging	–
TOTAL	40.0000		10.0000	**TOTAL**	35.0000		15.0000

5.2.2 Distributed Energy Management Scheduling Analysis

Based on the above simulation system, we carry out two optimal scheduling simulations according to the load forecasting value. The virtual electricity-power load demand are 40 and 35 MW; the virtual thermal-power load demand are 10 and 15 MW, which have been predicted by the load forecasting algorithm mentioned above. The parameters of electricity/thermal power outputs and initial value can be seen in **Table 4**.

According to the above virtual load, based on the proposed distributed economic optimal scheduling algorithm, the optimal scheduling scheme of each piece of equipment under the power balance constraints, ship navigation constraints, and equipment safety constraints [i.e., (8)–(11)] is solved which satisfies the electricity/thermal load demand and the constraints of BESS. In this example, in order to meet the electricity/thermal load demand, the relevant generation units need to adjust their output power to cope with the change of load demand. **Figure 7** is the trajectory of the electricity-power outputs error, and the thermal-power outputs error is shown in **Figure 8**. According to the above two figures, it is noted that after 60 optimization iterations, the total power outputs of the system satisfy the first virtual load scheduling goal and after 100 optimization iterations, the system meets the second virtual load scheduling goal.

The electricity-power and thermal-power outputs trajectories can be seen in **Figures 9**, **10**. According to the curves of total power outputs, it is noted that the optimization scheme analyzed by the proposed algorithm can deal with the energy management problem based on the S-IES, effectively. The power outputs have little fluctuation, which can meet the demand of electricity load and thermal load in a timely manner. The optimal scheduling scheme based on twice virtual electricity/thermal load energy management can be seen in **Table 5**.

According to the above results, BESS is in the charging state during the twice optimal scheduling, and the total power outputs of renewable energy equipment, i.e., 4-PVs, are significantly higher than that of traditional fuel energy supply equipment, i.e., 3CHPs. Therefore, it can be proved that the optimal scheduling scheme based on the optimization method proposed in this paper can effectively improve the environmental benefits of the ship-integrated energy system, which can reduce the greenhouse gas emission during the sailing voyage as well.

It can be seen from the above results that under the reasonable scheduling strategy, the S-IES could achieve the coordination and optimization of multiple energy sources, and significantly improve the flexibility and economic efficiency of the energy system on an intelligent ship.

6 CONCLUSION

In this paper, an integrated energy system in an intelligent ship, considering the characteristics of the electricity-power system and thermal-power system, has been constructed, which improves the utilization efficiency of energy and reduces greenhouse gas emissions. Simultaneously, this paper presents a short-term load forecasting method based on an ensemble learning algorithm, which can quickly and accurately predict the load demand during the whole voyage. In addition, for ensuring reliable and stable navigation, this paper proposes a distributed optimal scheduling scheme, which can deal with the problem of energy management in the S-IES suitably. It is noted that the distributed scheduling scheme proposed in this paper can obtain the social benefits and ensure safe sailing synchronously, which can be proved by the simulation results. In future, for promoting the intellectualization of the maritime industry, volatility and disturbance factors will be further analyzed.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/Supplementary Material, further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

QS designed the experiments, research methods, and the format analysis. The tools analysis, data processing, and writing the original draft were carried out by YZ. FT performed the writing-review on references. TL contributed to proof reading

and project/organization management. All authors have read and agreed to the published version of the manuscript.

FUNDING

This work is supported in part by the National Natural Science Foundation of China (under grant nos. 51939001, 61803064, 61751202, 61903092, 61976033, and U1813203); the Science and Technology Innovation Funds of Dalian (under grant no. 2018J11CY022); the Liaoning Revitalization Talents Program (under grant nos. XLYC1908018 and XLYC1807046); the Natural Science Foundation of Liaoning (2019-ZD-0151, 20170540098); the Fundamental Research Funds for the Central Universities (under grant nos. 3132019345, 3132020103, and 3132020125).

REFERENCES

Balsamo, F., De Falco, P., Mottola, F., and Pagano, M. (2020). Power Flow Approach for Modeling Shipboard Power System in Presence of Energy Storage and Energy Management Systems. *IEEE Trans. Energ. Convers.* 35 (4), 1944–1953. doi:10.1109/TEC.2020.2997307

Chen, L., Tong, Y., and Dong, Z. (2020). Li-Ion Battery Performance Degradation Modeling for the Optimal Design and Energy Management of Electrified Propulsion Systems. *Energies* 13, 1629. doi:10.3390/en13071629

Czermański, E., Pawłowska, B., Oniszczuk-Jastrzabek, A., and Cirella, G. T. (2020). Decarbonization of Maritime Transport: Analysis of External Costs. *Front. Energ. Res.* 8, 28. doi:10.3389/fenrg.2020.00028

Falsone, A., Notarnicola, I., Notarstefano, G., and Prandini, M. (2020). Tracking-ADMM for Distributed Constraint-Coupled Optimization. *Automatica* 117, 108962. doi:10.1016/j.automatica.2020.108962

Fang, S., Xu, Y., Li, Z., Ding, Z., Liu, L., and Wang, H. (2019). Optimal Sizing of Shipboard Carbon Capture System for Maritime Greenhouse Emission Control. *IEEE Trans. Ind. Appl.* 55 (6), 5543–5553. doi:10.1109/TIA.2019.2934088

Fang, S., Xu, Y., Wen, S., Zhao, T., Wang, H., and Liu, L. (2020). Data-Driven Robust Coordination of Generation and Demand-Side in Photovoltaic Integrated All-Electric Ship Microgrids. *IEEE Trans. Power Syst.* 35 (3), 1783–1795. doi:10.1109/TPWRS.2019.2954676

Farrok, O., Islam Sheikh, M. R., Guo, Y., Zhu, J., and Lei, G. (2018). Oceanic Wave Energy Conversion by a Novel Permanent Magnet Linear Generator Capable of Preventing Demagnetization. *IEEE Trans. Industry Appl.* 54 (6), 6005–6014. doi:10.1109/TIA.2018.2863661

Feng, W., Huang, W., and Ren, J. (2018). Class Imbalance Ensemble Learning Based on the Margin Theory. *Appl. Sci.* 8, 815. doi:10.3390/app8050815

Gangatharan, S., Rengasamy, M., Elavarasan, R. M., Das, N., Hossain, E., and Sundaram, V. M. (2020). A Novel Battery Supported Energy Management System for the Effective Handling of Feeble Power in Hybrid Microgrid Environment. *IEEE Access* 8, 217391–217415. doi:10.1109/ACCESS.2020.3039403

Kanellos, F. D., Tsekouras, G. J., and Hatziargyriou, N. D. (2014). Optimal Demand-Side Management and Power Generation Scheduling in an All-Electric Ship. *IEEE Trans. Sustainable Energ.* 5 (4), 1166–1175. doi:10.1109/TSTE.2014.2336973

Kanellos, F. D., Anvari-Moghaddam, A., and Guerrero, J. M. (2016). Smart Shipboard Power System Operation and Management. *Inventions* 1, 22. doi:10.3390/inventions1040022

Kanellos, F. D. (2014). Optimal Power Management with GHG Emissions Limitation in All-Electric Ship Power Systems Comprising Energy Storage Systems. *IEEE Trans. Power Syst.* 29 (1), 330–339. doi:10.1109/TPWRS.2013.2280064

Lei, L., Gao, T., Liu, Y., and Tong, S. (2021). Time-varying IBLFs-Based Adaptive Control of Uncertain Nonlinear Systems with Full State Constraints. *Automatica* 129, 109595. doi:10.1016/j.automatica.2021.109595

Li, Y., Gao, W., Zhang, H., and Zhou, J. (2020a). A Distributed Double-Newton Descent Algorithm for Cooperative Energy Management of Multiple Energy Bodies in Energy Internet. *IEEE Trans. Ind. Inform.* 17 (9), 5993–6003. doi:10.1109/TII.2020.3029974

Li, Y., Gao, D. W., Gao, W., Zhang, H., and Zhou, J. (2020b). Double-Mode Energy Management for Multi-Energy System via Distributed Dynamic Event-Triggered Newton-Raphson Algorithm. *IEEE Trans. Smart Grid* 11 (6), 5339–5356. doi:10.1109/TSG.2020.3005179

Li, J., Zhang, G., Liu, C., and Zhang, W. (2020). COLREGs-Constrained Adaptive Fuzzy Event-Triggered Control for Underactuated Surface Vessels with the Actuator Failures. *IEEE Trans. Fuzzy Syst.* doi:10.1109/TFUZZ.2020.3028907

Liang, H., Zhang, Y., Huang, T., and Ma, H. (2020). Prescribed Performance Cooperative Control for Multi-Agent Systems with Unknown Control Directions and Input Quantization. *IEEE Trans. Cybern.* 50 (5), 1810–1819. doi:10.1109/TCYB.2019.2893645

Liu, Z., Saberi, A., Stoorvogel, A. A., and Nojavanzadeh, D. (2020). H∞ Almost State Synchronization for Homogeneous Networks of Non-introspective Agents: A Scale-free Protocol Design. *Automatica* 122, 109276. doi:10.1016/j.automatica.2020.109276

Long, H., Xu, S., Lu, X., Yang, Z., Li, C., Jing, J., et al. (2020). Data-Driven Hybrid Equivalent Dynamic Modeling of Multiple Photovoltaic Power Stations Based on Ensemble Gated Recurrent Unit. *Front. Energ. Res.* 8, 185. doi:10.3389/fenrg.2020.00185

Ma, D., Hu, X., Zhang, H., Sun, Q., and Xie, X. (2021). A Hierarchical Event Detection Method Based on Spectral Theory of Multidimensional Matrix for Power System. *IEEE Trans. Syst. Man, Cybern. Syst.* 51 (4), 2173–2186. doi:10.1109/TSMC.2019.2931316

Peng, Z., and Wang, D. (2018). Distributed Maneuvering of Autonomous Surface Vehicles Based on Neurodynamic Optimization and Fuzzy Approximation. *IEEE Trans. Control. Syst. Technol.* 26 (3), 1083–1090. doi:10.1109/TCST.2017.2699167

Quan, H., Srinivasan, D., and Khosravi, K. (2014). Short-Term Load and Wind Power Forecasting Using Neural Network-Based Prediction Intervals. *IEEE Trans. Neural Networks Learn. Syst.* 25 (2), 303–315. doi:10.1109/TNNLS.2013.2276053

Rafiei, M., Boudjadar, J., and Khooban, M. -H. (2020). Energy Management of a Zero-Emission Ferry Boat with a Fuel-Cell-Based Hybrid Energy System: Feasibility Assessment. *IEEE Trans. Ind. Electron.* 68 (2), 1739–1748. doi:10.1109/TIE.2020.2992005

Sun, Q., Fan, R., Li, Y., and Huang, B., (2019a). A Distributed Double-Consensus Algorithm for Residential We-Energy. *IEEE Trans. Ind. Inform.* 15 (8), 4830–4842. doi:10.1109/TII.2019.2921431

Sun, Q., Zhang, N., You, S., and Wang, J. (2019b). The Dual Control with Consideration of Security Operation and Economic Efficiency for Energy Hub. *IEEE Trans. Smart Grid* 10 (6), 5930–5941. doi:10.1109/TSG.2019.2893285

Teng, F., Shan, Q., and Li, T. (2020). Intelligent Ship Integrated Energy System and its Distributed Optimal Scheduling Algorithm. *Acta Autom. Sin.* 46, 1809–1817. doi:10.16383/j.aas.c200176

Wang, R., Sun, Q., Ma, D., and Hu, X. (2020). Line Impedance Cooperative Stability Region Identification Method for Grid-Tied Inverters under Weak Grids. *IEEE Trans. Smart Grid* 11 (4), 2856–2866. doi:10.1109/TSG.2020.2970174

Wang, J., Wang, Z., Chen, X., and Qiu, J. (2021). Synchronization Criteria of Delayed Inertial Neural Networks with Generally Markovian Jumping. *Neural Networks* 139, 64–76. doi:10.1016/j.neunet.2021.02.004

Wen, S., Zhao, T., Tang, Y., Xu, Y., Zhu, M., and Huang, Y. (2020). A Joint Photovoltaic-dependent Navigation Routing and Energy Storage System Sizing Scheme for More Efficient All-Electric Ships. *IEEE Trans. Transp. Electrification* 6 (3), 1279–1289. doi:10.1109/TTE.2020.3015983

Wen, S., Zhao, T., Tang, Y., Xu, Y., Zhu, M., Fang, S., et al. (2021). Coordinated Optimal Energy Management and Voyage Scheduling for All-Electric Ships Based on Predicted Shore-Side Electricity Price. *Trans. Industry Appl.* 57 (1), 139–148. doi:10.1109/TIA.2020.3034290

Ye, P., Wang, T., and Wang, F. (2018). A Survey of Cognitive Architectures in the Past 20 Years. *IEEE Trans. Cybernetics* 48 (12), 3280–3290. doi:10.1109/TCYB.2018.2857704

Zhang, H., Li, Y., Gao, D. W., and Zhou, J. (2017). Distributed Optimal Energy Management for Energy Internet. *IEEE Trans. Ind. Inform.* 13 (6), 3081–3097. doi:10.1109/TII.2017.2714199

Zhang, G., Yao, M., Xu, J., and Zhang, W. (2020). Robust Neural Event-Triggered Control for Dynamic Positioning Ships with Actuator Faults. *Ocean Eng.* 207, 107292. doi:10.1016/j.oceaneng.2020.107292

Zhang, J., Chen, X., and Gu, G. (2020). State Consensus for Discrete-Time Multi-Agent Systems over Time-Varying Graphs. *IEEE Trans. Automatic Control.* 66 (1), 346–353. doi:10.1109/TAC.2020.2979750

Zhang, Y., Shan, Q., Li, T., and Teng, F. (2020). "Energy Dispatch Scheme on Ship Integrated Energy System with Photovoalatic and CHP," in 2020 Chinese Automation Congress (CAC), Shanghai, China, November 6–8, 2020, 3339–3344. doi:10.1109/CAC51589.2020.9326598

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Energy Research

ORIGINAL RESEARCH
published: 21 September 2021
doi: 10.3389/fenrg.2021.733910

An Adaptive Sliding Mode Control Based on Disturbance Observer for LFC

Mofan Wei[1,2], *Sheng Lin*[1,2]*, *Yan Zhao*[1,2], *Hao Wang*[3] and *Qian Liu*[4]

[1]School of Renewable Energy, Shenyang Institute of Engineering, Shenyang, China, [2]Key Laboratory of Regional Multi-energy System Integration and Control of Liaoning Province, Shenyang, China, [3]School of Electrical Engineering, Shenyang University of Technology, Shenyang, China, [4]Liaoning Province Information Centre, Shenyang, China

In the power system, the loads and nonlinearity parameters cause the system frequency deviation, which complicates the load frequency control (LFC). To deal with the above problem, an adaptive sliding mode control (SMC) based on disturbance observer is proposed to eliminate frequency deviation for interconnected power system in this paper. Firstly, the mathematical model of the power system is established, where the power exchange between the tie line is considered as the variable of the designed sliding surface. Secondly, the nonlinear disturbance observer is constructed to estimate the parameter uncertainty and load of power system. Thirdly, combined with the estimated value of the disturbance observer and integral sliding mode surface, the SMC is designed. Moreover, considering the inherent shortcoming of SMC—the chattering problem, an adaptive strategy is applied to the SMC to ensure the stability of controller. Next, the stability of the designed SMC is proved by Lyapunov stability theory. Finally, to verify the effectiveness of the proposed controller, several simulations are presented.

Keywords: load frequency control, interconnected power system, the disturbance observer, the adaptive control, sliding mode control

OPEN ACCESS

Edited by:
Dazhong Ma,
Northeastern University, China

Reviewed by:
Jingwei Hu,
Northeastern University, China
Guangliang Liu,
Bohai University, China

***Correspondence:**
Sheng Lin
linsheng_sie@163.com

Specialty section:
This article was submitted to Smart Grids, a section of the journal Frontiers in Energy Research

Received: *30 June 2021*
Accepted: *07 September 2021*
Published: *21 September 2021*

Citation:
Wei M, Lin S, Zhao Y, Wang H and Liu Q (2021) An Adaptive Sliding Mode Control Based on Disturbance Observer for LFC. Front. Energy Res. 9:733910. doi: 10.3389/fenrg.2021.733910

INTRODUCTION

LFC is a crucial technology for stable operation of modern large-scale interconnected power systems. Due to random uncertainty such as the power demand, the power generation, the communication time lag, the device parameter, etc. The frequency and power exchange in a large range power system will fluctuate or escape from the planned tolerance. When unexpected uncertainty occurs in power system, the purpose of frequency control is to quickly stabilize the system frequency and exchange power between interconnected systems within an acceptable plan (Dou et al., 2017). Driven by the rapid development of technology, power generation units, electrical equipment, and communication systems, power systems have become more complex (Wang et al., 2021b). Therefore, an effective frequency control strategy urgently needs to be proposed to manage the challenge of complex systems.

Furthermore, micro-grid can fully exploit renewable energy to reduce carbon emission. Based on the advantages of the micro-grid, the system has been widely established. However, when micro-grid encounters the intermittency of renewable resource, the rapid fluctuations of load and the uncertainties of internal parameters, frequency regulation is more complex (Kahrobaeian and Mohamed, 2012; Khooban et al., 2017; Lu et al., 2017). To stabilize the frequency, many mature control algorithms have been implemented to solve LFC problems, such as intelligent control,

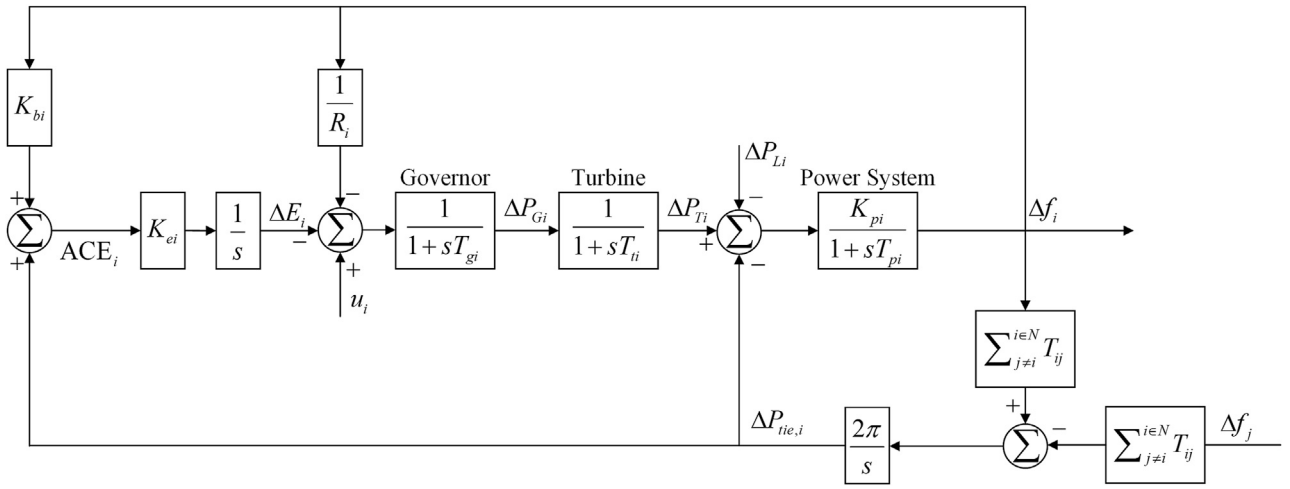

FIGURE 1 | Block diagram of interconnected system.

adaptive control (Rashidi et al., 2004), robust control (Huang et al., 2016; Jiang et al., 2012), fuzzy control (Yousef et al., 2014), proportional-integral differentiation (PID) control (Khodabakhshian and Edrisi, 2008; Tan, 2010), etc.

PID control is the common control tool to damp frequency oscillation for micro-grid which is treated as a linear model (Bevrani and Hiyama, 2008; Kamwa et al., 2001). The PID controllers are the simple and easy control tool which can powerfully tuned for several specific operation points (Wangdee and Billinton, 2006). As the integration of renewable energy power generation, the characteristics of power system are non-linear, thus the PID control has no ability to eliminate frequency deviation, especially when the actual work point of micro-grid deviates far from the expected work point (Tang et al., 2015). In (Farahani et al., 2012), the PID controller were optimized to eliminate frequency deviation. The main idea is to tune the gains of PID controller by the lozi map-based chaotic algorithm. Thus, a scheduling PID control strategy based on optimized parameters was applied to microgrid. Similarly, T. Chaiyatham proposed the fuzzy logic-PID controller which utilizes bee colony optimization to tune the fuzzy logic-PID controllers of micro-grid (Chaiyatham et al., 2019). For non-reheat thermal system, Gonggui Chen et al. utilized the fuzzy PID controller based on the Improved Ant Colony Optimization algorithm against system frequency deviation (Chen et al., 2020). Lim et al. (Lim et al., 1998) solved the LFC problem for the unmeasurable state in micro-grid using the robust control theory. Meanwhile, the adaptive control schemes were validated with system parameter uncertainties (Pan and Liaw, 1989).

Generally, as a well-known control method, SMC is a non-linear variable structure controller, whose control is discontinuity. As the advantages of strong robustness properties and quick response, it has been studied extensively (Li et al., 2017; Li et al., 2018; Ai-Hamouz and AbdeI-Magid, 1993; Wang et al., 2021a; Mi et al., 2013). Therefore, SMC is an alternative control strategy to address load frequency problems in micro-grid. Due to external disturbances of renewable energy and parameter variations in internal system, SMC has gained significant research attentions in the field of LFC. AI Hamouz designed a variable structure controller using linear sliding surface, where a step disturbance was considered in the simulation (Ai-Hamouz and AbdeI-Magid, 1993). Therefore, there is a problem with this control strategy. In Sivaramakrishnan et al. (1984), a SMC was presented through the pole assignment technique, where the parametric uncertainties were studied with the matched condition. However, when the power grid is working normally, the system parameters cannot always meet the matched condition. Yang Mi proposed SMC based on proportion-lintegral sliding mode surface, and this method was proved that micro-grid can be immune to the external disturbances with matched and unmatched conditions. However, the aforementioned control strategy was studied based on the proportional-integral controller which has overshoot (Mi et al., 2013). In recent years, considering the parameter uncertainty and the complexity of external disturbances, the advanced SMC method based on disturbance observer was proposed (Li et al., 2016; Mu et al., 2017; Wang et al., 2019; Ma et al., 2017; Liu et al., 2016; Ginoya et al., 2014). In Mu et al. (2017), SMC with neural network observer was constructed, where the measured values were used to control law and it was proven to be superior in the simulation. However, the uncertainties of system parameters are not demonstrated. Ark Dev proposed a SMC based on Luenberger observer (Dev and Sarkar, 2019). However, the Luenberger observer is applied in the absence of rigorous theoretical proof.

In this paper, a disturbance observer, which estimates the matched and unmatched disturbances in the power system, is applied in the LFC. Furthermore, an adaptive SMC strategy based on disturbance observer is investigated to eliminate frequency deviation.

The main contributions of this paper are described as.

1) The disturbance observers are proposed and applied to estimate the disturbances of the multi-area interconnected

power system, which effectively track matching and unmatching disturbances.

2) SMC is improved to eliminate frequency deviation. Firstly, comparing the traditional proportional-integral sliding mode surface, the area control error (ACE) and estimated value are taken as state variables into the novel sliding mode surface, which ensures that the frequency deviation and the ACE converge to the equilibrium point. Secondly, to address the chattering of controller, the adaptive law is designed.

MODEL OF POWER SYSTEM

In the power system, the frequency deviation is caused by the fluctuation of the load. The function of LFC is to eliminate frequency deviation. The system proposed in many documents has been applied to LFC. In this chapter, the mathematical model of the power system is established where the types of disturbances are elaborated.

LFC block diagram of i th area power system is illustrated in **Figure 1**. Due to the complexity of the power system structure, it is regarded as a nonlinear system in practice. However, since the load fluctuation is very small, linearized power system model is approved for theoretical analysis. In this section, N multi-region interconnected systems that connect subsystems through tie lines are studied. When the system is disturbed, the system is adjusted by primary frequency control, which can restore the system frequency to the planned tolerance. Then, SMC is adopted to eliminate frequency deviation.

The mathematical dynamics of N regional systems can be expressed as

$$\begin{aligned}
\Delta\dot{f}_i(t) &= -\frac{1}{T_{pi}}\Delta f_i(t) + \frac{K_{pi}}{T_{pi}}\Delta P_{Ti}(t) - \frac{K_{pi}}{T_{pi}}\Delta P_{tie,i}(t) - \frac{K_{pi}}{T_{pi}}\Delta P_{Li}(t)\\
\Delta\dot{P}_{Ti}(t) &= -\frac{1}{T_{ti}}\Delta P_{Ti}(t) + \frac{1}{T_{ti}}\Delta P_{Gi}(t)\\
\Delta\dot{E}_i(t) &= K_{ei}K_{bi}\Delta f_i(t) + K_{ei}\Delta P_{tie,i}(t)\\
\Delta\dot{P}_{tie,i}(t) &= 2\pi\sum_{j\neq i}^{i\in N} T_{ij}\Delta f_i(t) - 2\pi\sum_{j\neq i}^{i\in N} T_{ij}\Delta f_j(t)\\
\Delta\dot{P}_{Gi}(t) &= -\frac{1}{R_i T_{gi}}\Delta f_i(t) - \frac{1}{T_{gi}}\Delta E_i(t) + \frac{1}{T_{gi}}u_i(t)
\end{aligned} \tag{1}$$

where $\Delta f_i(t)$, $\Delta P_{Ti}(t)$, $\Delta E_i(t)$, $\Delta P_{tie,i}(t)$, $\Delta P_{Gi}(t)$ are the deviation of frequency, machine mechanical output, integral control, tie-line power, valve position, respectively; T_{pi}, T_{ti}, T_{gi} are power system time constants, turbine time constants, governor time constants, respectively; K_{ei}, K_{bi} and K_{pi} are integral control gain, frequency bias factor and power system gain, respectively; R_i is speed regulation coefficient; T_{ij} is the tie-line co-efficient between area i and j ; $i = 1, 2, 3, \cdots, N$ and N represents the number of subsystems.

In this paper, the i th ACE can be expressed as

$$\mathrm{ACE}_i = \Delta P_{tie,i}(t) + K_{bi}\Delta f_i(t) \tag{2}$$

Based on **Eq. 1**, the matrix form of power system can be expressed as

$$\dot{x}_i(t) = A_i x_i(t) + B_i u_i(t) + E_{ij}x_j(t) + L_i\Delta P_{Li}(t) \tag{3}$$

where

$$x_i = \begin{bmatrix} \Delta f_i(t) & \Delta P_{Ti}(t) & \Delta E_i(t) & \Delta P_{tie,i}(t) & \Delta P_{Gi}(t) \end{bmatrix}$$

$$A_i = \begin{bmatrix} -\frac{1}{T_{pi}} & \frac{K_{pi}}{T_{pi}} & 0 & -\frac{K_{pi}}{T_{pi}} & 0\\ 0 & -\frac{1}{T_{ti}} & 0 & 0 & \frac{1}{T_{ti}}\\ K_{ei}K_{bi} & 0 & 0 & K_{ei} & 0\\ 2\pi\sum_{j\neq i}^{i\in N} T_{ij} & 0 & 0 & 0 & 0\\ -\frac{1}{R_i T_{gi}} & 0 & -\frac{1}{T_{gi}} & 0 & 0 \end{bmatrix}$$

$$B_i = \begin{bmatrix} 0 & 0 & 0 & 0 & \frac{1}{T_{gi}} \end{bmatrix}^T,$$

$$E_{ij} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 0 & 0\\ -2\pi\sum_{j\neq i}^{i\in N} T_{ij} & 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$L_i = \begin{bmatrix} -\frac{K_{pi}}{T_{pi}} & 0 & 0 & 0 & 0 \end{bmatrix}^T$$

where A_i is the system matrix; B_i is the input matrix; E_{ij} is the interconnection matrix; L_i is the load disturbance matrix.

When the system is working, there is uncertainty in the system matrix, which is an important factor leading to frequency deviation. In this section, the uncertainty of system parameters and load disturbances are regarded as concentrated disturbances which are characterized by linearity and nonlinearity. Then, according to **Eq. 3**, we get

$$\dot{x}_i(t) = A_i x_i(t) + B_i u_i(t) + E_{ij}x_j(t) + \Gamma_i(t) \tag{4}$$

$$\Gamma_i(t) = A_i^{\vee} x_i(t) + B_i^{\vee} u_i(t) + E_{ij}^{\vee} x_j(t) + L_i\Delta P_{Li}(t) \tag{5}$$

where $A_i^{\vee}$, $B_i^{\vee}$ and $E_{ij}^{\vee}$ are matrices with uncertain parameters. Furthermore, assume that the integrated disturbance is matched/mismatched and bounded.

$$\|\Gamma_i(t)\| \leq \alpha.$$

where α is a positive constant.

DESIGN OF SMC WITH DISTURBANCE OBSERVER

Design of Disturbance Observer

In the power system, there are matching and mismatching disturbances, which are unknown. The uncertainty and load satisfy the following assumption:

Assumption 1. $rank[B_i, \Gamma_i(t)] = rank[B_i]$ or $rank[B_i, \Gamma_i(t)] \neq rank[B_i]$.

Based on Assumption 1, the concentrated disturbance can be expressed as follows

$$\Gamma_i(t) = \begin{bmatrix} d_{i1} & d_{i2} & d_{i3} & d_{i4} & d_{i5} \end{bmatrix}^T$$

With the concentrated disturbance composed of system parameter uncertainty and load, a nonlinear disturbance observer is designed to estimate the unknown disturbance as follows

$$\begin{bmatrix} \hat{d}_{i1} \\ \hat{d}_{i2} \\ \hat{d}_{i3} \\ \hat{d}_{i4} \\ \hat{d}_{i5} \end{bmatrix} = \begin{bmatrix} p_{i1} \\ p_{i2} \\ p_{i3} \\ p_{i4} \\ p_{i5} \end{bmatrix} + L_i x_i \tag{6}$$

$$\begin{bmatrix} \dot{p}_{i1} \\ \dot{p}_{i2} \\ \dot{p}_{i3} \\ \dot{p}_{i4} \\ \dot{p}_{i5} \end{bmatrix} = -L_i \left(L_i x_i + \begin{pmatrix} p_{i1} \\ p_{i2} \\ p_{i3} \\ p_{i4} \\ p_{i5} \end{pmatrix} \right) - L_i \left(A_i x_i(t) + B_i u_i(t) + E_{ij} x_j(t) \right) \tag{7}$$

where $\hat{\Gamma}_i(t) = \begin{bmatrix} \hat{d}_{i1} & \hat{d}_{i2} & \hat{d}_{i3} & \hat{d}_{i4} & \hat{d}_{i5} \end{bmatrix}^T$ is the estimated value of the disturbance, which is the matched and unmatched disturbance of the system. L_i is the designed observer matrix gain, $\begin{bmatrix} p_{i1} & p_{i2} & p_{i3} & p_{i4} & p_{i5} \end{bmatrix}^T$ is the auxiliary vector.

Design of SMC

SMC has been proven to be a powerful controller in many documents, and it is adopted to many fields, such as aircraft, robots, and inverted pendulums. Traditionally, there are two crucial steps in SMC, which are the sliding surface and the control law. The sliding surface ensures that the system state reaches the equilibrium point, and the control law drives the state of the space to the sliding surface. For matching and matching disturbances, the proportional-integral sliding surface is generally adopted in SMC.

$$s(t) = C_{i1} x_i + C_{i2} \int_0^t x_i dt \tag{8}$$

where C_{i1} and C_{i2} are the vectors of design parameters. The dimensions of vectors are $C_{i1} \in \mathbb{R}^{6\times1}$ and $C_{i2} \in \mathbb{R}^{6\times1}$. In an interconnected power system, the sliding mode surface is constructed based on $\Delta f_i(t)$, $\Delta P_{Ti}(t)$, $\Delta E_i(t)$, $\Delta P_{tie,i}(t)$ and $\Delta P_{Gi}(t)$ to ensure that the system state converges in a finite time. *ACE*, calculated by the integral, cannot guarantee the adjustment to zero, which drives the frequency deviation to escape the scheduled scope. Based on the above analysis, we improved the sliding surface to meet the LFC of the interconnected system.

The improved sliding surface is

$$\hat{s}_i(t) = C_{i1} x_i + C_{i2} \int_0^t x_i dt + c_{i11} \left(\int_0^t ACE_i dt + \int_0^t \times \int_0^t ACE_i dt\, dt \right) + C_i \hat{\Gamma}_i(t) \tag{9}$$

where $C_{i1} = \begin{bmatrix} c_{i1} & c_{i2} & c_{i3} & c_{i4} & c_{i5} \end{bmatrix}$ and $C_{i2} = \begin{bmatrix} c_{i6} & c_{i7} & c_{i8} & c_{i9} & c_{i10} \end{bmatrix}$ are the designed parameters, and $C_i = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \end{bmatrix}$. c_{i11} is a positive constant.

In SMC, the chattering problem is difficult to address. In this paper, the adaptive control is used to slow down the output chattering of the controller. Based on **Eq. 9**, the adaptive controller can be obtained

$$u_i(t) = -(C_{i1}B_i)^{-1} \left(C_{i1} A_i x_i(t) + c_{i11} \dot{A}CE(t) + C_{i2} x_i + c_{i11} \int_0^t ACE dt + C_{i1} E_{ij} x_j(t) + C_{i1} \hat{\Gamma}_i(t) + \alpha_i \hat{s}(t) + \hat{\beta}_i sign\left(\hat{s}(t) \right) \right) \tag{10}$$

where c_{i11} and α_i are the positive constants, $sign(*)$ is the symbolic function, $\hat{\beta}_i$ is the adaptive control law. The definition of $\hat{\beta}_i$ is as follows:

$$\dot{\hat{\beta}}_i = k \left\| \hat{s}_i(t) \right\| \tag{11}$$

where k is a positive constant.

STABILITY ANALYSIS

In this section, the stability of disturbance observer and SMC is proved.

Stability Analysis of Disturbance Observer

To prove that the disturbance observer can track matched/unmatched disturbances, the following assumptions are necessary.

Assumption 2. The derivative of the disturbance in the system satisfies $\lim_{t\to\infty} \dot{\Gamma}_i(t) = 0$.

Assumption 3. The error e_i in the system is bounded, that is, $\|e_i\| \le \|e_i^*\|$.

$$e_i = \Gamma_i(t) - \hat{\Gamma}_i(t) \tag{12}$$

where e_i^* is a positive constant.

Proof:

Based on Assumption 3, the derivative of the error is given

$$\dot{e}_i = \dot{\Gamma}_i(t) - \dot{\hat{\Gamma}}_i(t) \tag{13}$$

Combining **Eqs 6**, **8**, **9**, we get

$$\begin{aligned} \dot{e}_i &= \dot{\Gamma}_i(t) - \begin{bmatrix} \dot{p}_{i1} & \dot{p}_{i2} & \dot{p}_{i3} & \dot{p}_{i4} & \dot{p}_{i5} \end{bmatrix}^T - L_i \dot{x}_i \\ &= \dot{\Gamma}_i(t) - \begin{bmatrix} \dot{p}_{i1} & \dot{p}_{i2} & \dot{p}_{i3} & \dot{p}_{i4} & \dot{p}_{i5} \end{bmatrix}^T \\ &\quad -L_i \left(A_i x_i(t) + B_i u_i(t) + E_{ij} x_j(t) + \Gamma_i(t) \right) \\ &= \dot{\Gamma}_i(t) - L_i \left(L_i x_i + \begin{bmatrix} p_{i1} & p_{i2} & p_{i3} & p_{i4} & p_{i5} \end{bmatrix}^T \right) \\ &\quad + L_i \left(A_i x_i(t) + B_i u_i(t) + E_{ij} x_j(t) \right) - L_i \left(A_i x_i(t) + B_i u_i(t) \right. \\ &\quad \left. + E_{ij} x_j(t) + \Gamma_i(t) \right) = \dot{\Gamma}_i(t) - L_i \left(\Gamma_i(t) - \hat{\Gamma}_i(t) \right) \le -L_i \left\| e_i^* \right\| \end{aligned} \tag{14}$$

By means of **Eq. 14**, it can be concluded that the observer can estimate the disturbance in a finite time.

Stability Analysis of Improved SMC

For an adaptive controller **Eq. 10**, it is necessary to prove the stability of system **Eq. 4**. Next, the Lyapunov approach is adopted to analysis the stability of the system under the controller.

Proof:

TABLE 1 | The Parameters of interconnection system.

Area	T_{pi}	T_{ti}	T_{gi}	K_{pi}	K_{ei}	K_{bi}	E_{ij}
1	20	0.3	0.08	120	10	0.41	$E_{12} = 0.5$
2	25	0.33	0.07	113	9	0.37	$E_{21} = 0.5$

The Lyapunov function is constructed as follows:

$$S_i = \frac{1}{2}\widehat{s}_i^2(t) + \frac{1}{2k}\widetilde{\beta}_i^2 \tag{15}$$

where $\widetilde{\beta}_i = \widehat{\beta} - \overline{\beta}$, $\dot{\widetilde{\beta}}_i = \dot{\widehat{\beta}}_i - \dot{\overline{\beta}}_i = \dot{\widehat{\beta}}_i$. $\overline{\beta}_i$ is a positive constant.

The derivative of S_i becomes

$$\dot{S}_i = \widehat{s}_i(t) \cdot \dot{\widehat{s}}_i(t) + \frac{1}{k}\widetilde{\beta}_i \cdot \dot{\widetilde{\beta}}_i \tag{16}$$

Differentiating improved sliding surface **Eq. 9**, we get

$$\dot{\widehat{s}}_i(t) = C_{i1}\dot{x} + C_{i2}x_i + c_{i11}\dot{A}CE_i(t) + c_{i11}\int_0^t ACE_i dt + C_i\dot{\widehat{\Gamma}}_i(t) \tag{17}$$

Substituting **Eq. 4**, we get

$$\begin{aligned} S_i = \widehat{s}_i(t) \cdot \Bigg(& C_{i1}\left(A_i x_i(t) + B_i u_i(t) + E_{ij}x_j(t) + \Gamma_i(t)\right) \\ & + C_{i2}x_i + c_{i11}\dot{A}CE_i(t) + c_{i11}\int_0^t ACE_i dt + C_i\dot{\widehat{\Gamma}}_i(t)\Bigg) + \frac{1}{k}\widetilde{\beta}_i \cdot \dot{\widetilde{\beta}}_i \end{aligned} \tag{18}$$

Using **Eqs 6**, **10**, **11**, the $\dot{S}$ is as

$$\begin{aligned} \dot{S} &= \widehat{s}_i(t)\left(C_{i1}\left(\Gamma_i(t) - \widehat{\Gamma}_i(t)\right) - \alpha_i\widehat{s}_i(t) - \widehat{\beta}_i sign\left(\widehat{s}_i(t)\right)\right) \\ &+ \frac{1}{k}\left(\widehat{\beta}_i - \overline{\beta}_i\right) \cdot \dot{\widehat{\beta}}_i \le -\alpha_i\left\|\widehat{s}_i(t)\right\|^2 - \widehat{\beta}_i \cdot \widehat{s}_i(t) \cdot sign\left(\widehat{s}_i(t)\right) \\ &- \frac{1}{k}\dot{\widehat{\beta}}_i \cdot \overline{\beta}_i + \widehat{\beta}_i\left\|\widehat{s}_i(t)\right\| = -\alpha_i\left\|\widehat{s}_i(t)\right\|^2 - \overline{\beta}_i\left\|\widehat{s}_i(t)\right\| \le 0 \end{aligned} \tag{19}$$

where $\overline{\beta}_i \ge \left\|C_{i1}e_i^*\right\|$.

From **Eq. 19**, we can conclude that when the coefficients of the controller are selected appropriately, the frequency deviation of system **Eq. 4** is eliminated with the controller **Eq. 10**.

SIMULATION ANALYSIS

Several simulations are presented for improved SMC in this section. In the simulation, frequency deviation of single and interconnected systems is analyzed. First, in presence of load fluctuation, the control performances are presented, such as, the frequency deviation, the sliding mode surface, and the controller output. Secondly, when there are parameter uncertainties and load disturbances in the interconnected system, the designed SMC performance is analyzed. The parameters of the system are shown in **Table 1** (Mi et al., 2013).

Single-Area Power System

The step load disturbances are applied to the system. The load disturbance is applied to the system, which is a −0.1 p.u. disturbance applied on the system at 0–10 s. Moreover, the

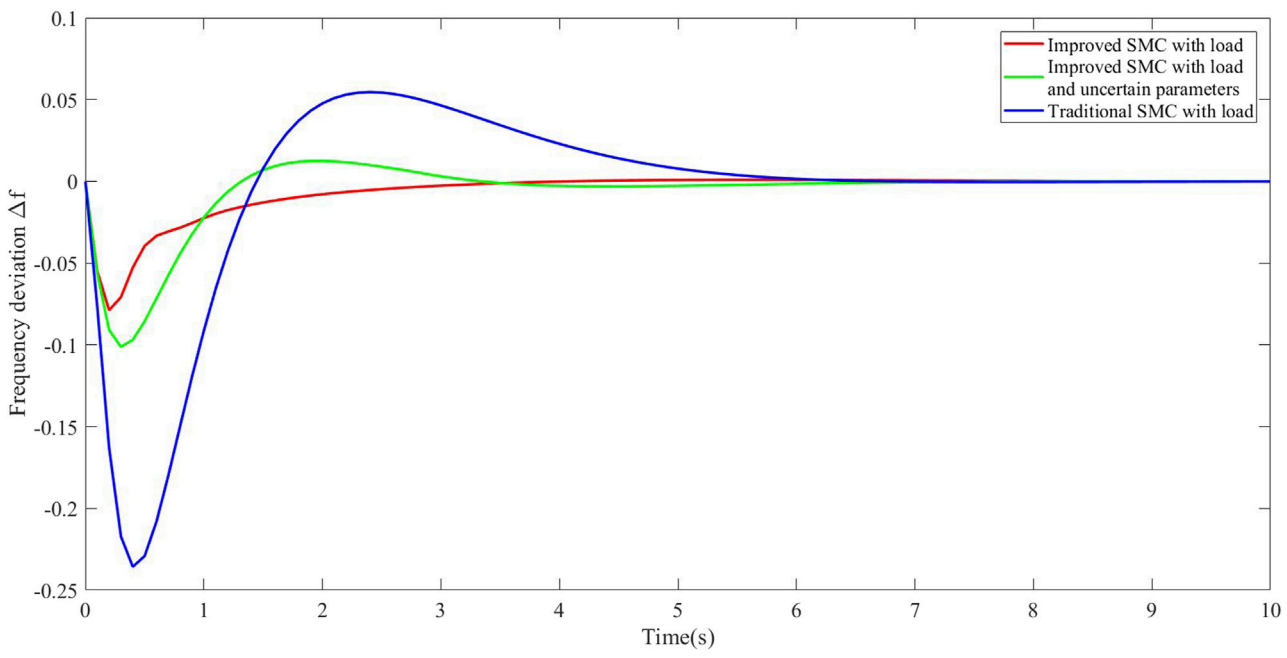

FIGURE 2 | The frequency deviation.

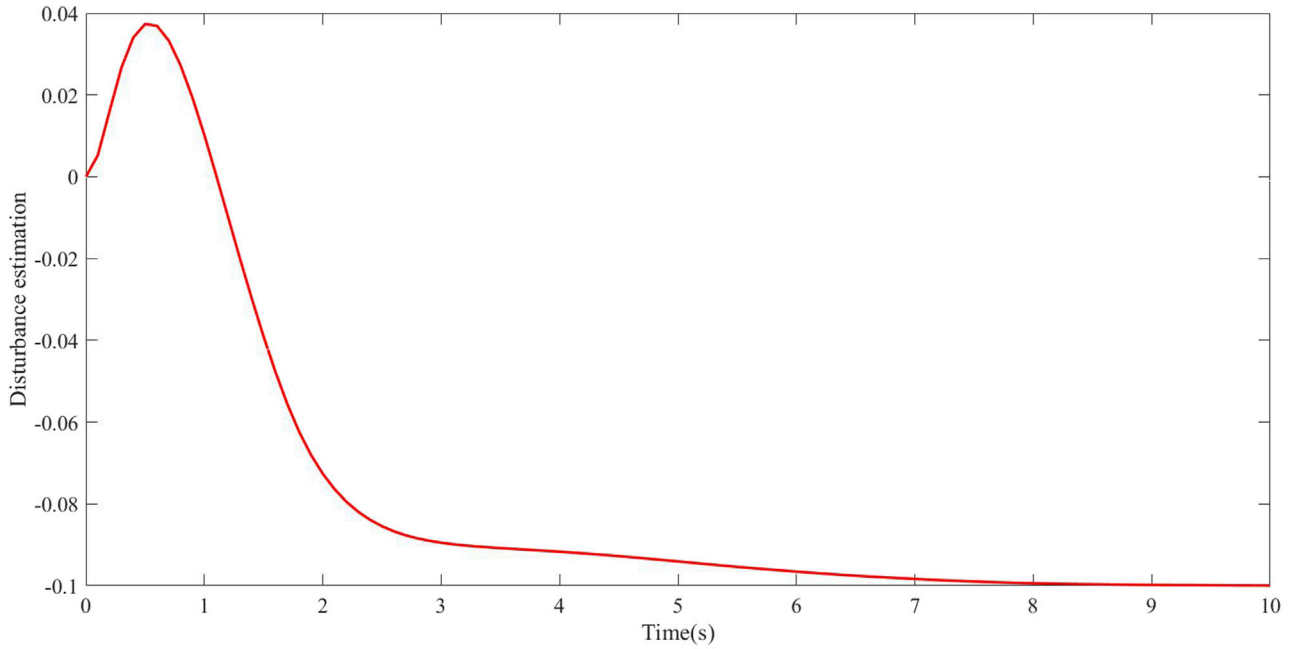

FIGURE 3 | The disturbance observation.

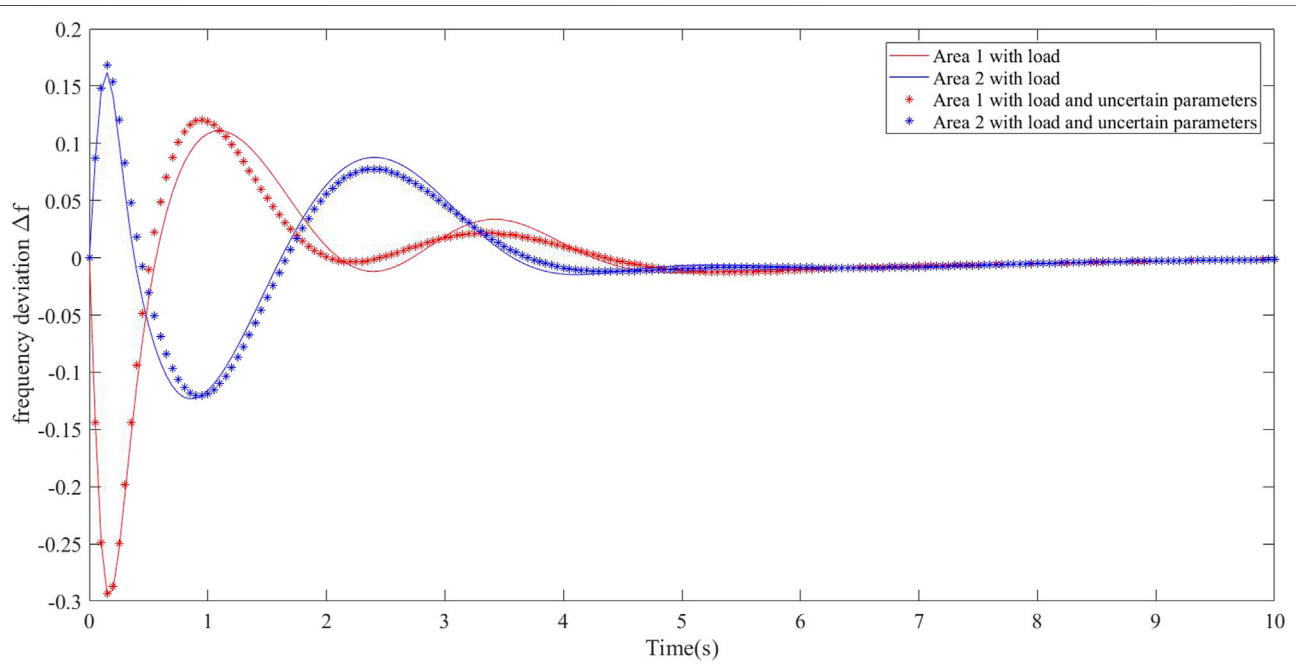

FIGURE 4 | Frequency deviation of interconnected systems.

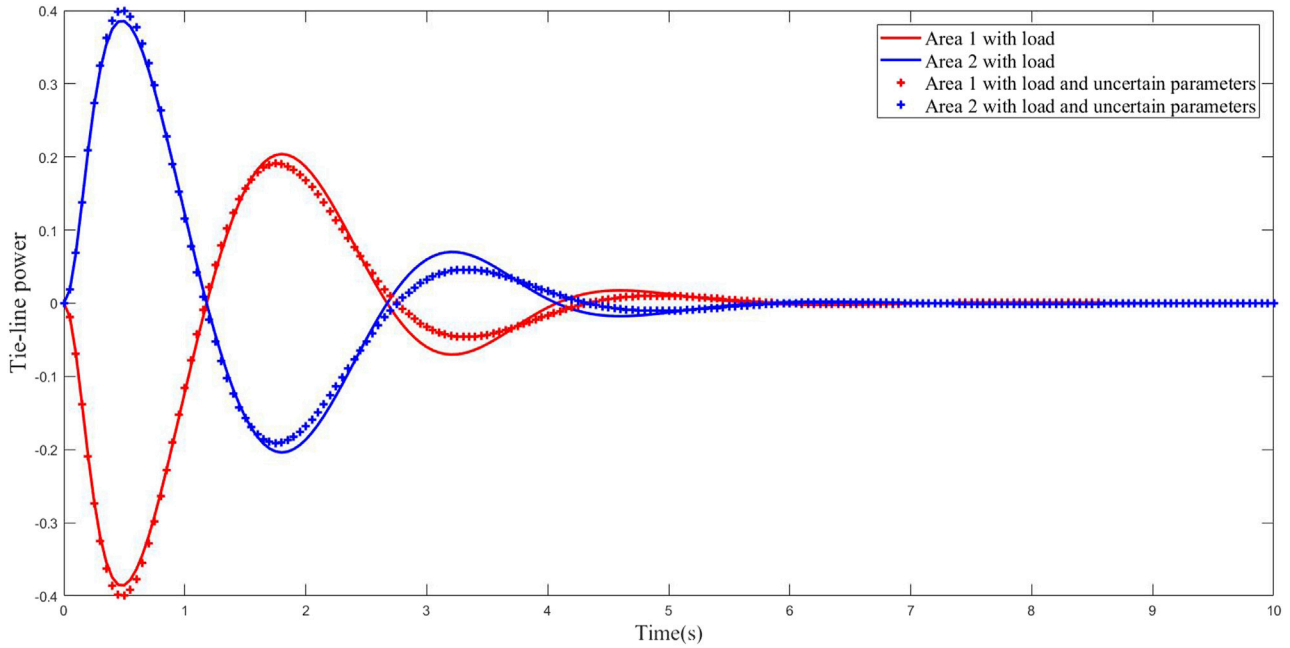

FIGURE 5 | The power exchange of tie line.

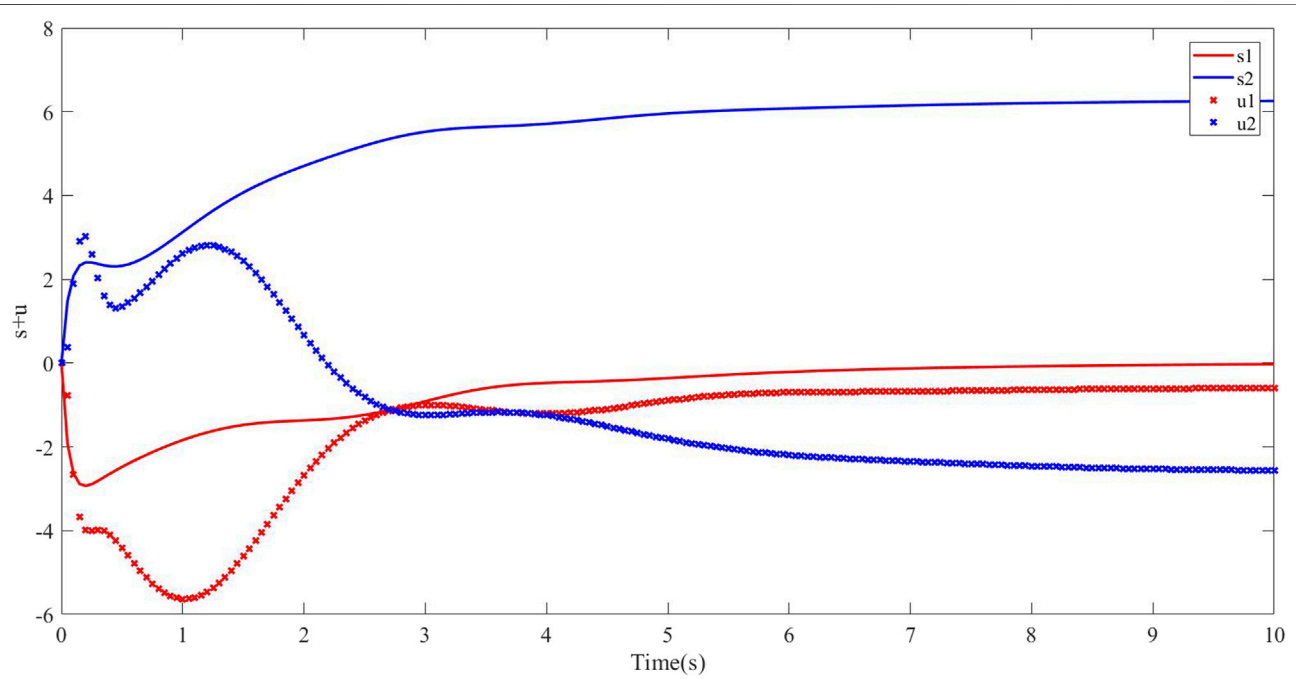

FIGURE 6 | Sliding mode surface and control law with multiple systems.

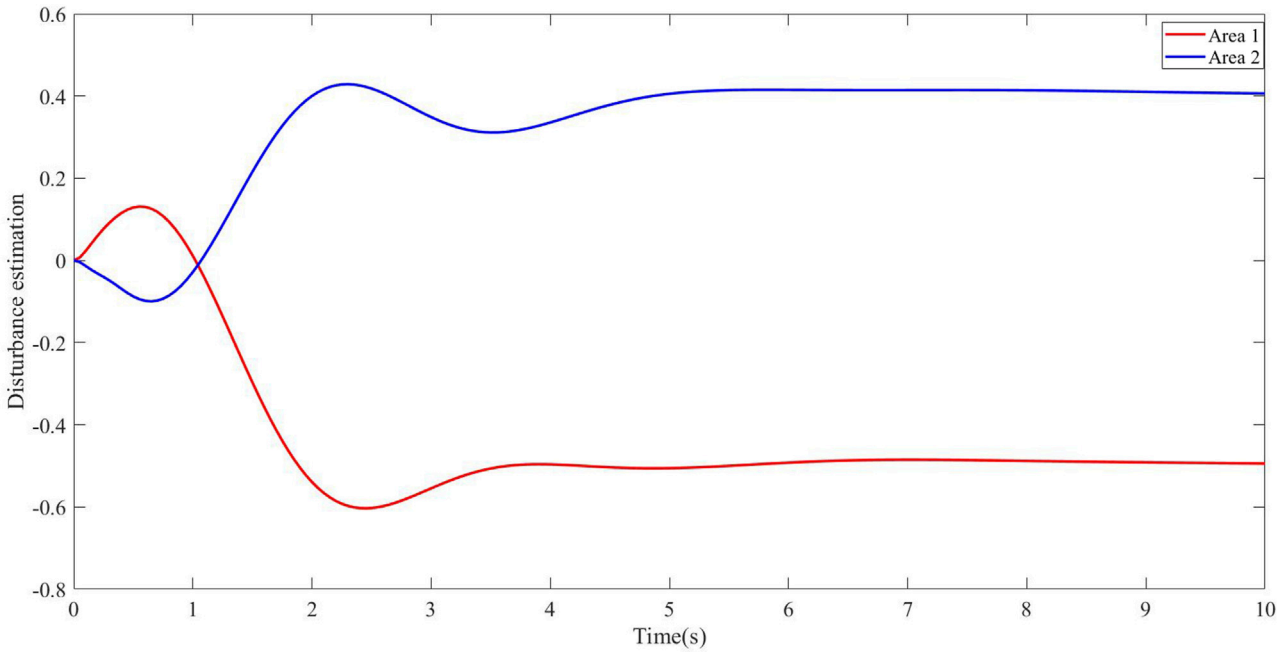

FIGURE 7 | Disturbance estimation.

parameter uncertainty in the single system is analyzed. The parameters of disturbance observer, SMC and parameter uncertainty are as follows:

$$C_{11} = \begin{bmatrix} 20 & 9 & 1 & 1 & 1 \end{bmatrix}^T, C_{12} = \begin{bmatrix} 23 & 8 & 4 & 3 & 2 \end{bmatrix}^T,$$

$$c_{111} = 1,\ \alpha_1 = 6,\ k = 10,\ L_1 = \begin{bmatrix} -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix},$$

$$A_1^{\vee} = \begin{bmatrix} 0 & e^{-t} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

The frequency deviation with traditional SMC and improved SMC is shown in **Figure 2**. Compared with the traditional SMC, the designed SMC effectively suppresses the frequency deviation. In addition, the control strategy has been verified to effectively eliminate the frequency deviation caused by the uncertainty of the system parameters. From **Figure 3**, it concludes that the designed disturbance observer can track the load disturbance.

Multi-Area Power System

There are linear disturbances and parameter uncertainties in multi-area systems ($N = 2$). The load disturbance in area 1 is a −0.5 p.u., and the load disturbance in area 2 is a 0.4 p.u. The parameters of disturbance observer, SMC and parameter uncertainties in multi-area are as follows:

$$C_{21} = \begin{bmatrix} 25 & 6 & 3 & 1 & 1 \end{bmatrix}^T, C_{22} = \begin{bmatrix} 15 & 8 & 1 & 1 & 1 \end{bmatrix}^T,$$

$$c_{211} = 1,\ \alpha_2 = 6,\ L_2 = L_1,\ A_1^{\vee} = \begin{bmatrix} 0 & 0 & 0 & \sin(t) & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix},$$

$$A_2^{\vee} = \begin{bmatrix} 0 & 0 & 0 & -\cos(t) & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

In multi-area systems, it can be concluded that when the system has disturbances and nonlinear parameter uncertainties, the system frequency can be eliminated with the designed SMC from **Figure 4**. In addition, we can know that the power exchange of tie line converges to zero at 10 s from **Figure 5**. The sliding mode surface and control law in the system are presented in **Figure 6**. In **Figure 7**, the estimated values of the disturbance observer in area 1 and area 2 can effectively estimate the load disturbance.

CONCLUSION

In this paper, the frequency in the power system, regarded as the most basic feature of the power system, is solved by the designed control strategy. An improved SMC is proposed, which guarantees the stability of the system with disturbances. Firstly, the

disturbance observer is used in LFC, which calculates the disturbance. Furthermore, it is proved by Lyapunov stability theory. Secondly, the adaptive SMC based on the disturbance observer is designed, which destroys the conservativeness of the traditional SMC. Then, it is proved to ensure the system stability. Finally, several simulation results are presented. In addition, for power systems with nonlinear characteristics, the advanced control strategy will be further studied.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

AUTHOR CONTRIBUTIONS

SL provided the main idea of this paper. MW and YZ were responsible for the writing work of this paper. HW and QL carried out the simulation.

FUNDING

This work was supported by Liaoning Revitalization Talents Program (XLYC1907138), Natural Science Foundation of Liaoning Province (2019-MS-239), Key R&D Program of Liaoning Province (2020JH2/10300101) and Technology Innovation Talent Fund of Shenyang (Grant Nos. RC190360, RC200252).

REFERENCES

Ai-Hamouz, Z. M., and AbdeI-Magid, Y. L. (1993). Variable Structure Load Frequency Controllers for Multiarea Power Systems. *Int. J. Electr. Power Energ. Syst.* 15 (5), 293–300.

Bevrani, H., and Hiyama, T. (2008). Robust Decentralised PI Based LFC Design for Time Delay Power Systems. *Energ. Convers. Management* 49 (2), 193–204. doi:10.1016/j.enconman.2007.06.021

Chaiyatham, T., Ngamroo, I., Pothiya, S., and Vachirasricirikul, S. (2019). Design of Optimal Fuzzy Logic-PID Controller Using Bee colony Optimization for Frequency Control in an Isolated Wind-Diesel System. 2009 Proceedings of the Transmission & Distribution Conference & Exposition: Asia and Pacific, Seoul Korea (South) 1–4. doi:10.1109/TD-ASIA.2009.5356804

Chen, G., Li, Z., Zhang, Z., and Li, S. (2020). An Improved ACO Algorithm Optimized Fuzzy PID Controller for Load Frequency Control in Multi Area Interconnected Power Systems. *IEEE Access* 8, 6429–6447. doi:10.1109/access.2019.2960380

Dev, A., and Sarkar, M. K. (2019). Robust Higher Order Observer Based Nonlinear Super Twisting Load Frequency Control for Multi Area Power Systems via Sliding Mode. *Int. J. Control. Autom. Syst.* 17 (7), 1814–1825. doi:10.1007/s12555-018-0529-4

Dou, C., Yue, D., Guerrero, J. M., Xie, X., and Hu, S. (2017). Multiagent System-Based Distributed Coordinated Control for Radial DC Microgrid Considering Transmission Time Delays. *IEEE Trans. Smart Grid* 8 (5), 2370–2381. doi:10.1109/tsg.2016.2524688

Farahani, M., Ganjefar, S., and Alizadeh, M. (2012). PID Controller Adjustment Using Chaotic Optimisation Algorithm for Multi-Area Load Frequency Control. *IET Control. Theor. Appl.* 6 (13), 1984–1992. doi:10.1049/iet-cta.2011.0405

Ginoya, D., Shendge, P. D., and Phadke, S. B. (2014). Sliding Mode Control for Mismatched Uncertain Systems Using an Extended Disturbance Observer. *IEEE Trans. Ind. Electron.* 61 (4), 1983–1992. doi:10.1109/tie.2013.2271597

Huang, C., Yue, D., Xie, X., and Xie, J. (2016). Anti-windup Load Frequency Controller Design for Multi-Area Power System with Generation Rate Constraint. *Energies* 9 (5), 1–18. doi:10.3390/en9050330

Jiang, L., Yao, W., Wu, Q. H., Wen, J. Y., and Cheng, S. J. (2012). Delay-dependent Stability for Load Frequency Control with Constant and Time-Varying Delays. *IEEE Trans. Power Syst.* 27 (2), 932–941. doi:10.1109/tpwrs.2011.2172821

Kahrobaeian, A., and Mohamed, Y. A.-R. I. (2012). Interactive Distributed Generation Interface for Flexible Micro-grid Operation in Smart Distribution Systems. *IEEE Trans. Sustain. Energ.* 3 (2), 295–305. doi:10.1109/tste.2011.2178045

Kamwa, I., Grondin, R., and Hebert, Y. (2001). Wide-area Measurement Based Stabilizing Control of Large Power Systems-A Decentralized/hierarchical Approach. *IEEE Trans. Power Syst.* 16 (1), 136–153. doi:10.1109/59.910791

Khodabakhshian, A., and Edrisi, M. (2008). A New Robust PID Load Frequency Controller. *Control. Eng. Pract.* 16 (9), 1069–1080. doi:10.1016/j.conengprac.2007.12.003

Khooban, M. H., Niknam, T., Blaabjerg, F., and Dragičević, T. (2017). A New Load Frequency Control Strategy for Micro-grids with Considering Electrical Vehicles. *Electric Power Syst. Res.* 143, 585–598. doi:10.1016/j.epsr.2016.10.057

Li, H., Shi, P., and Yao, D. (2017). Adaptive Sliding-Mode Control of Markov Jump Nonlinear Systems with Actuator Faults. *IEEE Trans. Automat. Contr.* 62 (4), 1933–1939. doi:10.1109/tac.2016.2588885

Li, H., Shi, P., Yao, D., and Wu, L. (2016). Observer-based Adaptive Sliding Mode Control for Nonlinear Markovian Jump Systems. *Automatica* 64, 133–142. doi:10.1016/j.automatica.2015.11.007

Li, H., Wang, J., Wu, L., Lam, H.-K., and Gao, Y. (2018). Optimal Guaranteed Cost Sliding-Mode Control of Interval Type-2 Fuzzy Time-Delay Systems. *IEEE Trans. Fuzzy Syst.* 26 (1), 246–257. doi:10.1109/tfuzz.2017.2648855

Lim, K. Y., Wang, Y., Guo, G., and Zhou, R. (1998). A New Decentralized Robust Controller Design for Multi-Area Load Frequency Control via Incomplete State Feedback. *Optimal Control. Appl. Methods* 19 (5), 1998. doi:10.1002/(sici)1099-1514(199809/10)19:5<345:aid-oca634>3.0.co;2-5

Liu, X., Kong, X., and Lee, K. Y. (2016). Distributed Model Predictive Control for Load Frequency Control with Dynamic Fuzzy Valve Position Modelling for Hydrothermal Power System. *IET Control. Theor. Appl.* 10 (4), 1653–1664. doi:10.1049/iet-cta.2015.1021

Lu, X., Yu, X., Lai, J., Guerrero, J. M., and Zhou, H. (2017). Distributed Secondary Voltage and Frequency Control for Islanded Microgrids with Uncertain Communication Links. *IEEE Trans. Ind. Inf.* 13 (2), 448–460. doi:10.1109/tii.2016.2603844

Ma, M., Zhang, C., Liu, X., and Chen, H. (2017). Distributed Model Predictive Load Frequency Control of the Multi-Area Power System after Deregulation. *IEEE Trans. Ind. Electron.* 64 (6), 5129–5139. doi:10.1109/tie.2016.2613923

Mi, Y., Fu, Y., Wang, C., and Wang, P. (2013). Decentralized Sliding Mode Load Frequency Control for Multi-Area Power Systems. *IEEE Trans. Power Syst.* 28 (4), 4301–4309. doi:10.1109/tpwrs.2013.2277131

Mu, C., Liu, W., Xu, W., and Rabiul Islam, M. (2017). Observer-based Load Frequency Control for Island Microgrid with Photovoltaic Power. *Int. J. Photoenergy* 2017, 1–11. doi:10.1155/2017/2851436

Pan, C. T., and Liaw, C. M. (1989). An Adaptive Controller for Power System Load-Frequency Control. *IEEE Trans. Power Syst.* 4 (1), 122–128. doi:10.1109/59.32469

Rashidi, M., Rashidi, F., Arjomand, A. S., and Sahragard, J. (2004). Design of a Robust and Adaptive Load Frequency Controller for Multi-Area Power Networks with System Parametric Un-certainties Using TDMLP Neural Network. IEEE International Conference on Systems, Man and Cybernetics. 3698–3703. The Hague, Netherlands. 10-13 Oct. 2004.

Sivaramakrishnan, A. Y., Hariharan, M. V., and Srisailam, M. C. (1984). Design of Variable-Structure Load-Frequency Controller Using Pole Assignment Technique. *Int. J. Control* 40 (3), 487–498. doi:10.1080/00207178408933289

Tang, Y., Yang, J., Yan, J., and He, H. (2015). Intelligent Load Frequency Controller Using GrADP for Island Smart Grid with Electric Vehicles and Renewable Resources. *Neurocomputing* 170, 406–416.

Utkin, V. I. (1992). *Sliding Modes in Control Optimization*. Berlin, Germany: Springer.

Wang, R., Sun, Q., Hu, W., Li, Y., Ma, D., and Wang, P. (2021a). SoC-based Droop Coefficients Stability Region Analysis of the Battery for Stand-Alone Supply Systems with Constant Power Loads. *IEEE Trans. Power Electron.* 36 (7), 7866–7879. doi:10.1109/tpel.2021.3049241

Wang, R., Sun, Q., Ma, D., and Liu, Z. (2019). The Small-Signal Stability Analysis of the Droop-Controlled Converter in Electromagnetic Timescale. *IEEE Trans. Sustain. Energ.* 10 (3), 1459–1469. doi:10.1109/tste.2019.2894633

Wang, R., Sun, Q., Tu, P., Xiao, J., Gui, Y., and Wang, P. (2021b). Reduced-order Aggregate Model for Large-Scale Converters with Inhomogeneous Initial Conditions in Dc Microgrids. *IEEE Trans. Energ. Convers.* 36 (3), 2473–2484. doi:10.1109/tec.2021.3050434

Wangdee, W., and Billinton, R. (2006). Considering Load-Carrying Capability and Wind Speed Correlation of WECS in Generation Adequacy Assessment. *IEEE Trans. Energ. Convers.* 21 (3), 734–741. doi:10.1109/tec.2006.875475

Wen Tan, W. (2010). Unified Tuning of PID Load Frequency Controller for Power Systems via IMC. *IEEE Trans. Power Syst.* 25 (1), 341–350. doi:10.1109/tpwrs.2009.2036463

Yousef, H. A., AL-Kharusi, K., Albadi, M. H., and Hosseinzadeh, N. (2014). Load Frequency Control of a Multi-Area Power System: an Adaptive Fuzzy Logic Approach. *IEEE Trans. Power Syst.* 29 (4), 1822–1830. doi:10.1109/tpwrs.2013.2297432

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Energy Research

ORIGINAL RESEARCH
published: 08 October 2021
doi: 10.3389/fenrg.2021.745744

Improved Genetic Algorithm and XGBoost Classifier for Power Transformer Fault Diagnosis

Zhanhong Wu[1], Mingbiao Zhou[2], Zhenheng Lin[1], Xuejun Chen[1] and Yonghua Huang[1]*

[1]Department of Electrical Engineering, School of Mechatronics and Information Engineering, Putian University, Putian, China, [2]State Grid Sanming Electric Power Supply Company, Sanming, China

Power transformer is an essential component for the stable and reliable operation of electrical power grid. The traditional transformer fault diagnostic methods based on dissolved gas analysis are limited due to the low accuracy of fault identification. In this study, an effective transformer fault diagnosis system is proposed to improve identification accuracy. The proposed approach combines an improved genetic algorithm (IGA) with the XGBoost to form a hybrid diagnosis network. The combination of the improved genetic algorithm and the XGBoost (IGA-XGBoost) forms the basic unit of the proposed method, which decomposes and reconstructs the transformer fault recognition problem into several minor problems IGA-XGBoosts can solve. The results of simulation experiments show that the IGA performs excellently in the combined optimization of input feature selection and the XGBoost parameter, and the proposed method can accurately identify the transformer fault types with an average accuracy of 99.2%. Compared to IEC ratios, dual triangle, support vector machine and common vector approach the diagnostic accuracy of the proposed method is improved by 30.2, 47.2, 11.2, and 3.6%, respectively. The proposed method can be a potential solution to identify the transformer fault types.

Keywords: transformer fault diagnosis, genetic algorithm, XGBoost, dissolved gas analysis, performance measures

OPEN ACCESS

Edited by:
Lei Xi,
China Three Gorges University, China

Reviewed by:
Narottam Das,
Central Queensland University, Australia
Xiaoshun Zhang,
Shantou University, China

****Correspondence:***
Zhanhong Wu
zhanhong.wu@outlook.com

Specialty section:
This article was submitted to Smart Grids, a section of the journal Frontiers in Energy Research

Received: *22 July 2021*
Accepted: *21 September 2021*
Published: *08 October 2021*

Citation:
Wu Z, Zhou M, Lin Z, Chen X and Huang Y (2021) Improved Genetic Algorithm and XGBoost Classifier for Power Transformer Fault Diagnosis. Front. Energy Res. 9:745744. doi: 10.3389/fenrg.2021.745744

1 INTRODUCTION

Power transformers are one of the most expensive, complex, and momentous equipment in electrical power systems. The faults of any power transformer online could cause considerable damage to the power system and lead to the interruption of the power supply. Therefore, the early detection of faults in transformers is vital to improving the reliability of the power system. Suffered the electrical and thermal stress during the operation, the transformer oil and organic insulating inside the transformer will be decomposed and generate different gases. Commonly, these dissolved gases include hydrogen (H_2), methane (CH_4), acetylene (C_2H_2), ethylene (C_2H_4), ethane (C_2H_6) and can provide abundant information about the internal states of the transformer. Based on the gas chromatography methods, the composition of the dissolved gases can be qualitatively and quantitatively measured and then used for the identification of the latent fault. There are three main kinds of chromatographic analysis method (Cheng and Yu, 2018) of dissolved gases namely the characteristic gas method (Fu et al., 2012), the gas production rate method (Nogami et al., 1995; Xi Chen et al., 2010; Zeng et al., 2011), and the three-ratio method (Jiang et al., 2014; Dhote and Helonde, 2012; Liu et al., 2020). The above methods generally utilize the concentration of a specific gaseous molecule or the ratios of several different molecules indicate the state of a power transformer (Shang et al., 2019). In addition, several

improved methods have been proposed and applied for transformer fault diagnosis, including the Roger method (Ghoneim et al., 2016), the basic triangular diagram (Singh and Bandyopadhyay, 2010), the dual triangle method (Shang et al., 2019), etc. However, these methods have their inherent shortcomings. For example, most of these traditional diagnosis methods only make a limited contribution to a transformer's fault diagnosis due to low diagnostic accuracy (Yadaiah and Ravi, 2011). Meanwhile, the three-ratio and improved three-ratio methods have disadvantages of incomplete coding and excessively absolute coding boundary (Cheng and Yu, 2018). Therefore, due to these defects of traditional methods, it is necessary to investigate new transformer fault diagnosis methods.

With the rapid development of computer science and artificial intelligence algorithms, many models are conducted by combining intelligence techniques with DGA methods to accurately detect fault types. The utilization of artificial neural network (ANN) (Colorado et al., 2011; Bhalla et al., 2012; Yi et al., 2016; Meng et al., 2010; Castro and Miranda, 2005; Miranda and Castro, 2005; Souahlia et al., 2012), expert system (Lin et al., 1993; Wang et al., 2000; Saha and Purkait, 2004; Mani and Jerome, 2014; Li et al., 2009), fuzzy theory (Huang et al., 1997; Mofizul Islam et al., 2000; Zhou et al., 1997; Fan et al., 2017; Naresh et al., 2008), grey system (Dong et al., 2003; Cheng et al., 2018), support vector machine (SVM) (Fei and Zhang, 2009; Fei et al., 2009; Liu et al., 2016; Niu Wu et al., 2010; Yin et al., 2011) and other theories have significantly improved the accuracy of fault identification. However, deficiencies occur together with these intelligent diagnostic approaches. Based on the ANN method, the intelligent fault diagnostic method is susceptible to be overfitting and may get a local optimum (Yuan et al., 2019). As for the expert system, the accuracy of this diagnostic method depends on the completeness of the expert knowledge, and this method cannot learn from new data samples automatically (Weigen Chen et al., 2009). In addition, fuzzy theory depends exceedingly on the experience of the researcher and is difficult to acquire an appropriate relationship between the input and output variables (Žarković and Stojković, 2017). SVM is originally a binary classification algorithm which makes it difficult to determine the parameters for multi-classification problem (Zhu et al., 2018). A single intelligent approach for transformer fault diagnosis has various shortcomings and can not reflect all the operation status of the transformer. Various intelligent algorithms can be combined to form a hybrid network for mutual complementation to solve complex problems, which has been applied in electricity. Researchers in (Xi et al., 2020) proposed a deep-reinforcement-learning-based three-network double-delay actor-critic (TDAC) control strategy for the automatic generation control (AGC) to deal with the strong random disturbance issues caused by renewable energy. Researchers in (Zhang et al., 2020) proposed a predictive control (MPC) based model combined with real-time optimal mileage based dispatch (OMD) for generating company responding to AGC dispatch signals in real-time. The above hybrid networks perform excellently in dealing with complex problems. As for the transformer fault diagnosis, a diagnostic method can be conducted with a hybrid network that combines different algorithms.

TABLE 1 | The candidate feature set for intelligent fault diagnosis methods. Gas concentrations (ppm) and ratio of gas concentrations in the table.

Feature set					
H_2	CH_4	C_2H_6	C_2H_4	C_2H_2	CH_4/H_2
C_2H_4/C_2H_6	C_2H_2/C_2H_4	C_2H_2/C_2H_6	TH	CH_4/TH	C_2H_4/TH
C_2H_2/TH	$TH1$	$CH_4/TH1$	$C_2H_6/TH1$	$C_2H_4/TH1$	$C_2H_2/TH1$

To achieve the objective of improving the accuracy of transformer fault diagnosis, a machine learning algorithm named XGBoost was employed as the classifier for the transformer fault identification in this paper, which is a scalable end-to-end tree boosting system (Chen and Guestrin, 2016). An improved genetic algorithm (IGA) is used for input feature selection and the XGBoost's optimization. Then an intelligent diagnostic method based on the combination of the IGA and the XGBoost classifier (IGA-XGBoost) is built. The remainder of this paper is organized as follows. **Section 2** presents the details of the proposed method, and **section 3** shows the experimental results and performance analysis. **Section 4** is the conclusion of this paper.

2 PROPOSED METHODS

In this section, the proposed method for power transformer faults detection and recognition is explained in detail. Different methods based on the artificial intelligence algorithms and DGA methods have been proposed to classify transformer faults, and the most significant issue which impacts the accuracy of fault classification is the appropriate selection of input features and classifiers (Tightiz et al., 2020). Consideration has been given to these two aspects in the proposed method.

2.1 Candidate Input Features

Intelligent transformer faults diagnosis methods proposed by other researchers commonly combine DGA methods with artificial intelligence algorithms in the last decades. The gas ratios or gas concentrations used in DGA methods are adopted as the inputs of these intelligent fault diagnosis methods. Nonetheless, not all the gas ratios or gas concentrations have the same significance for fault identification. Using uninformative features as inputs leads to artificial noise and poor performance in transformer faults diagnosis. Hence, effective features should be selected as the input, and uninformative features must be removed. In this study, following the traditional DGA methods, the concentrations of the dissolved gases or the ratios of several different gases are collected as candidate feature set for the input feature selection, as shown in **Table 1**. In **Table 1**, $TH = CH_4 + C_2H_4 + C_2H_2$ and $TH1 = CH_4 + C_2H_4 + C_2H_2 + C_2H_6$.

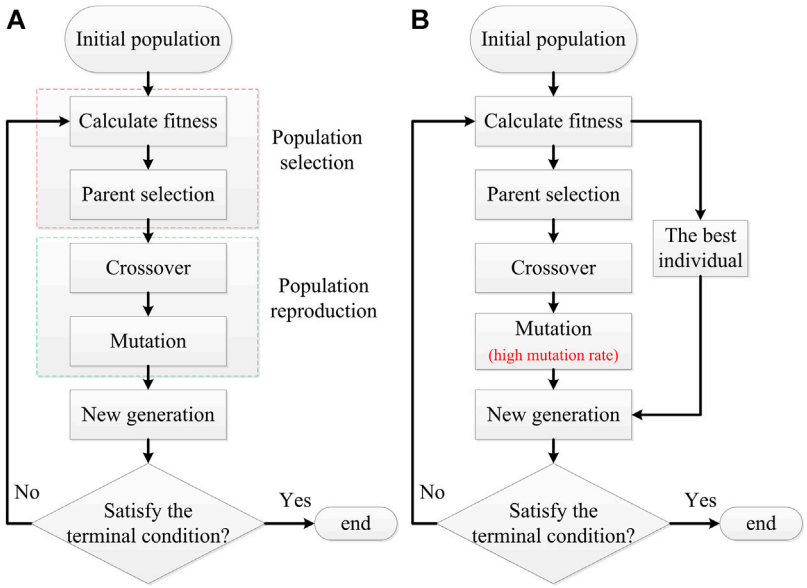

FIGURE 1 | The structures of the TGA and the IGA, **(A)** is TGA's structure, **(B)** is IGA's structure. The IGA differs from GA with two modifications: the high mutation rate of 0.3, and the other is the addition of elitist selection in the population selection process.

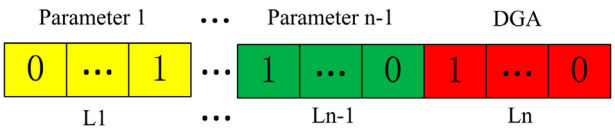

FIGURE 2 | Structure of chromosome coding. Chromosome coding contains n bit strings, of which L1 to Ln-1 represent the classifier's parameters, and the Ln bit string represents the DGA data used for the input feature.

2.2 Genetic Algorithm

2.2.1 Tradtional Genetic Algorithm and Improved Genetic Algorithm

It is common practice to separate the process of input feature selection from the classifier optimization process, which neglects interaction between the feature selection and the classifier optimization may lead to unreliable results (Daelemans et al., 2003). Combined optimization of the feature selection and the classifier' parameters can be achieved by genetic algorithm within a single approach. Since the traditional genetic algorithm (TGA) is prone to get trapped in the local optimal and fails to find the optimal global solution. In this paper, some improvements have been made to the TGA to enhance its global search capability, and an IGA is obtained. The IGA is utilized to combine the feature selection process with the classifier optimization process and assess which combination of input features and classifier's parameters substantially impact the accuracy of fault diagnosis to gain the optimal input features and the classifier's parameters. The difference between the TGA and the IGA is shown in **Figure 1**. **Figure 1A** shows the structure of the TGA, and **Figure 1B** shows the structure of the IGA. From 1A, it can be found that the main processes of the TGA include the population selection process and the population reproduction process. Two modifications make the IGA differ from the TGA: the high mutation rate of 0.3, and the other is the addition of elitist selection in the population selection process.

In the population selection process of the TGA, the generation of candidate solutions after initialization is called the population. Each individual of the generated population has its chromosome coding to represent the parameters of the classifier and the input features extracted from DGA data, as shown in **Figure 2**. Each chromosome coding contains n bit strings, of which L1 to Ln-1 represent the classifier's parameters, and the Ln bit string is used for the input feature selection. For feature selection, the bit with the value "1" in the Ln bit string represents the corresponding DGA feature is selected, and "0" means no selection. For the parameters setting of the classifier, the bit strings of parameters would be converted from the binary value to decimal value with a specific range by **Eq. 1**.

$$p = min_p + \frac{max_p - min_p}{2^l - 1} \times d \tag{1}$$

Here p represents chromosome coding of parameter, $\min_p$ represents minimum value of the parameter, $\max_p$ represents maximum value of the parameter, d represents decimal value of bit string, and l represents length of bit string.

The individuals of the population will then be selected for propagation by weighing their fitness values. The fitness values measure the population's performance. The fitness function described in **Eq. 2** uses the average accuracy of cross-validation for evaluation, and the higher the fitness value represents a better individual.

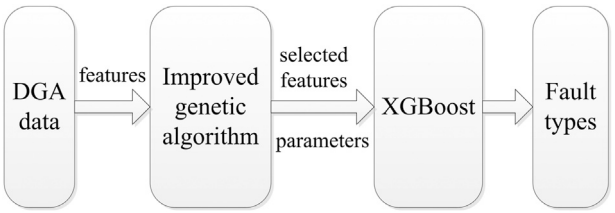

FIGURE 3 | Structure of the IGA-XGBoost. The IGA is used for the combined optimization of the input features and the XGBoost parameters, and the trained XGBoost is used to identify the fault types.

$$fitness = \frac{1}{k}\sum_{i=1}^{k} accuracy(i) \tag{2}$$

Here, k is the fold number of the cross-validation.

The probability of each individual of the population being selected is calculated in **Eq. 3** by the roulette wheel selection method.

$$probability(i) = \frac{fitness(i)}{\sum_{i=1}^{n} fitness(i)} \tag{3}$$

Here, n is the total number of individuals in each generation.

Then in the population reproduction process, crossover and mutation are employed to generate a new generation by the selected individuals of the population with a random mechanism. Crossover exchanges chromosome's segments between two selected individuals stochastically, and the bit value in the chromosome will be converted from "0" to "1" or vice versa occasionally in the mutation process. New individuals are formed through crossover and mutation, which are different from the original. A new generation is created in this way. Population reproduction and selection processes can be repeated under the "survival of the fittest" to achieve an optimal result.

In the roulette wheel selection method, the greater the individual's fitness is, the higher the possibility of such an individual would survive, but the optimal individual of each iteration still has a certain probability of being eliminated. Also, the process of crossover and mutation may lead to the disappearance of the optimal individual. Compared with TGA, the optimal individual of each iteration is added directly into the new generation in the IGA to avoid the disappearance of each iteration's optimal individual as shown in **Figure 1B**, which is called elitist selection. Besides, the mutation rate is set at 0.3 instead of the conventional low value to make the IGA jump out of the local optimum. These two modifications can effectively improve the global search capability of the IGA and the accuracy of transformer fault identification, as will be discussed in the Simulation Result.

2.2.2 Performance Measures

The main objective of the IGA is to enhance search capability for the optimization problem and gain better solutions. To analyze the enhancement of search capability, the following performance measures are defined (Sugihara, 1997).

1) Average fitness value $f(k)$: the average of the value obtained within k generations in n runs.

$$f(k) = \frac{\sum f_b(k)}{n} \tag{4}$$

Here, $f_b(k)$ is the best fitness values obtained within k generations; n is the number of independent runs.

2) Likelihood of evolution leap $Lel(k)$: the probability of average leaps within k generations among n independent runs. When a solution of one generation is better than the best solution obtained before the generation, the generation is said to be a leap.

$$Lel(k) = \frac{l}{n} \tag{5}$$

Here, l is the average number of leaps within k generations; n is the number of independent runs.

3) Likelihood of optimality $Lopt(k)$: the probability of obtaining optimal solutions within k generation in n independent runs.

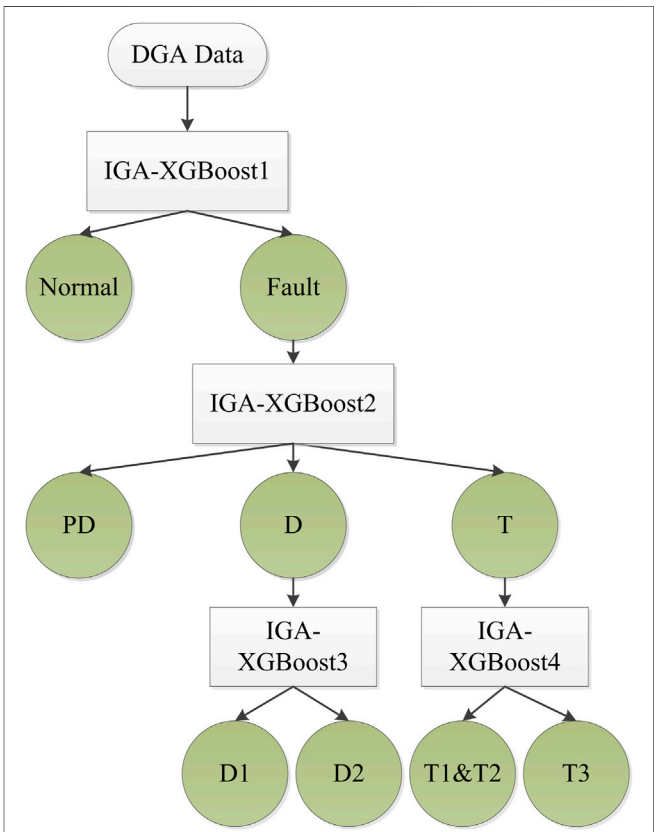

FIGURE 4 | Structure of the transformer fault diagnosis system based on the IGA-XGBoost. Transformer states are classified into six categories, and four IGA-XGBoost classifiers are used to diagnose the fault type step by step in the transformer fault diagnosis system.

TABLE 2 | Numbers of DGA data used in the training and test process.

Transformer states	Number of training samples	Number of test samples
NF	21	5
PD	16	3
D1	18	4
D2	23	6
T1&T2	23	3
T3	24	4
Total data sets	125	25

$$Lopt(k) = \frac{m}{n} \tag{6}$$

Here, m is the number of runs which produced an optimal solution within k generations; n is the number of independent runs.

2.3 Transformer Fault Diagnosis System

The XGBoost is a scalable tree boosting system that has been successfully applied in world-class machine learning and data mining competition because it is robust enough to avoid overfitting (Zhang and Zhan, 2017). In addition, the XGBoost algorithm can take advantage of the original data directly without normalization. Thus, the XGBoost is utilized as the classifier of the transformer fault diagnosis system in this study. The DGA data is not fed directly into the classifier in the proposed method. Since 18 features are collected from DGA methods, as shown in **Table 1**, the application of complete input data is too time-consuming and could lower the accuracy of faults classification due to the artificial noise. **Figure 3** shows the structure of the IGA-XGBoost. The IGA selects the input features fed to the XGBoost to decrease the input volume from 18 to a smaller number in the IGA-XGBoost. In addition, at the same time as input feature selection, the parameters of the XGBoost are decided by the IGA. The parameters of the XGBoost being decided by the IGA include eta, max_depth, min_child_weight, n_estimators, and n_gamma. The transformer fault diagnosis system described in **Figure 4** is developed based on the IGA-XGBoost. Transformer states are classified into six categories, which contain normal (N), partial discharge (PD), high-energy discharge (D1), low-energy discharge (D2), low and middle-temperature overheat (T1&T2), and high-temperature overheating (T3). The fault recognition problem is decomposed and reconstructed into several more minor problems that can be solved one by one. Four IGA-XGBoost classifiers are used to detect and identify transformer faults. The IGA-XGBoost1 is trained to separate the normal samples from the fault samples. The selected fault samples by the IGA-XGBoost1 are fed to the trained IGA-XGBoost2 and classified as PD, D, and T. Then, the IGA-XGBoost3 and the IGA-XGBoost4 are used to identify the D1, D2, T1&2, and T3.

3 SIMULATION RESULT

The DGA data set employed in this study is originated from Ref (Kirkbas et al., 2020). The data is divided into the training data set (125 samples) and the test data set (25 samples). These samples correspond to six states of the transformer. For each fault state, the number of samples used for the training and test process is shown in **Table 2**.

3.1 Performance of the Proposed Method

The proposed method is used for transformer fault diagnosis and compared with another transformer fault diagnosis system based on the TGA and the XGBoost, which has the same structure as the proposed method shown in **Figure 4**. The only difference between these two methods is that one uses the TGA while the other uses the IGA. To ensure the validity of the selected features and classifier parameters by IGA in the training process, the average accuracy of 8-fold cross-validation is taken as the fitness value. Therefore, the fitness curve is the average accuracy curve of cross-validation. The maximum generation number was 200. The initial population scale was set at 200, and the fitness of each iteration's best individual was collected to form the best fitness curve shown in **Figure 5**. **Figure 5** portrays the operation of the proposed method in five different independent implementations. **Figure 6** compares the proposed method and the transformer fault diagnosis system based on the TGA and the XGBoost for the global best fitness value in the training process.

It can be seen from **Figure 5** that the proposed method can achieve the same fitness value for the normal or fault (N-F) identification in different independent experiments using the IGA-XGBoost1. The global best fitness value can reach 99.22%. When detecting PD, D, or T fault (PD-D-T), the IGA-XGBoost2 also can gain the same high global fitness value of 99.04%. Besides, the proposed method can even 100% distinguish D1 and D2 faults (D1-D2) using the IGA-XGBoost3. Although the global fitness value was not as high as that of other faults identification when detecting T1&2 or T3 fault (T1&2-T3), most of them reached 97.92%, with only one fitness value reaching 95.83%. Compared with the method based on the traditional GA and the XGBoost, as can be seen from **Figure 6A**, the global fitness value of the method based on the traditional GA and the XGBoost varies with independent implementations, and its global best fitness value is also significantly lower than that of the proposed method (see **Figure 6B**). For N-F identification, the global best fitness value of the method based on the TGA ranged from 92.8 to 98.4%. As for T1&2-T3 identification, the global best fitness value of the method based on the TGA ranged from 86.5 to 98.1%. It can be seen from the above results that the IGA can achieve better solutions.

Performance measures such as average fitness value, likelihood of evolution leap, and likelihood of optimality have been taken into consideration to measure the enhancement of the IGA in the optimization problem. **Table 3** shows the average fitness value in the 100th and 200th generations for both the TGA and the IGA. **Table 4** shows the likelihood of evolution leap in the 100th and

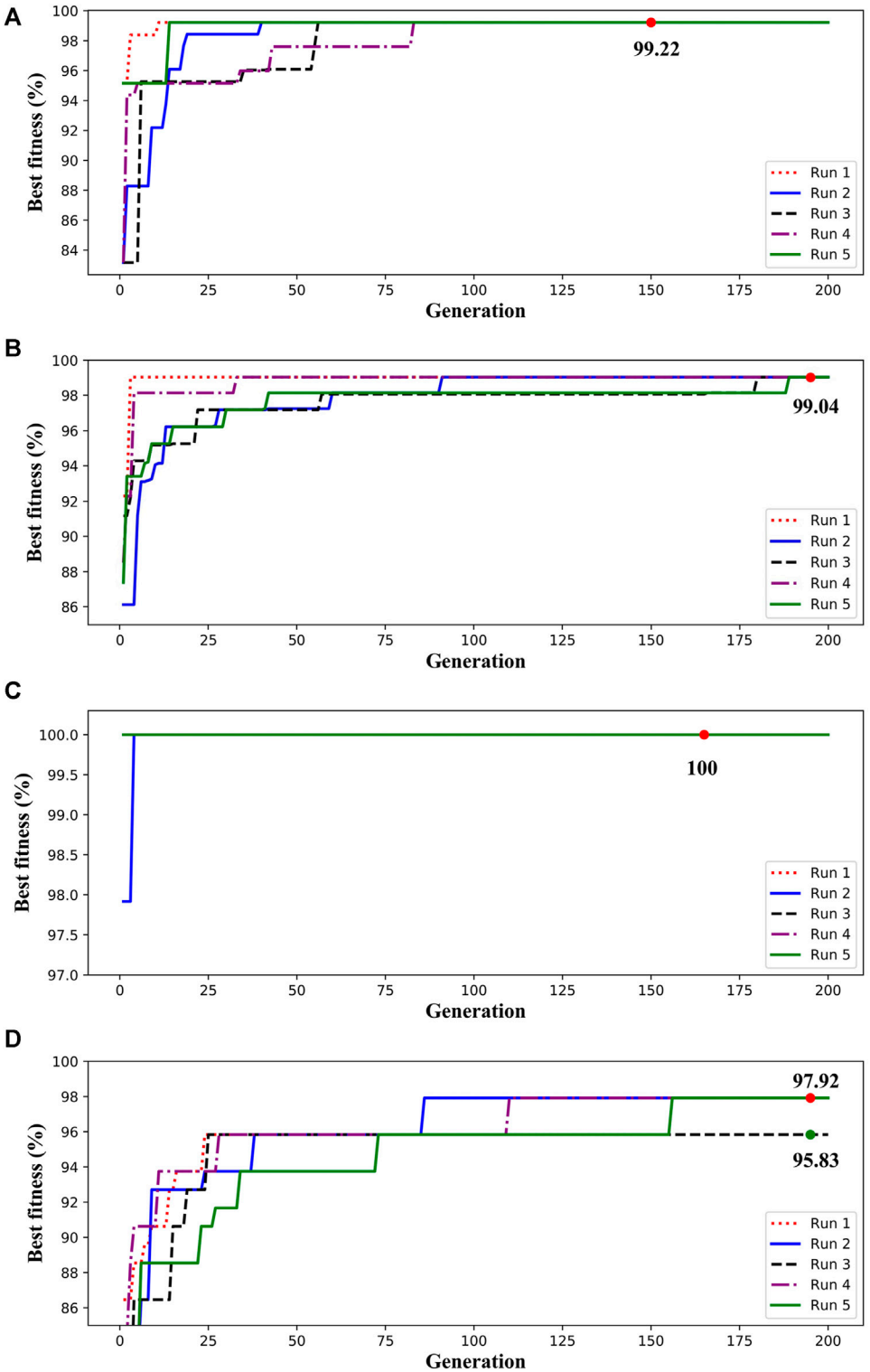

FIGURE 5 | Performance of the transformer fault diagnosis system in training process, **(A)** is the best fitness curves of the IGA-XGBoost1 for N-F identification, **(B)** is the best fitness curves of the IGA-XGBoost2 for PD-D-T identification, **(C)** is the best fitness curves of the IGA-XGBoost3 for D1-D2 identification, **(D)** is the best fitness curves of the IGA-XGBoost4 for T1&2-T3 identification.

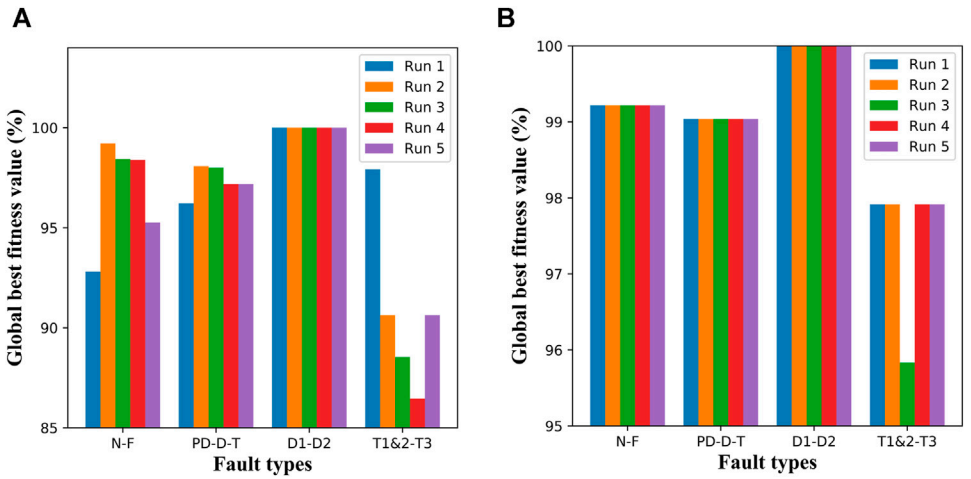

FIGURE 6 | The comparison of the global best fitness value among methods, **(A)** is the global best fitness value of the TGA, **(B)** is the global best fitness value of the IGA. The higher the global fitness value represents a better solution.

TABLE 3 | Average fitness value f(k).

Fault type	Traditional GA mutation rate = 0.01		Improved GA	
	After 100 generations	After 100 generations	After 100 generations	After 100 generations
N-F	0.968	0.968	0.992	0.992
PD-D-T	0.973	0.973	0.987	0.99
D1-D2	1	1	1	1
T1&2-T3	0.908	0.908	0.967	0.974

TABLE 4 | Likelihood of evolution leap Lel(k).

Fault type	Traditional GA mutation rate = 0.01		Improved GA	
	After 100 generations	After 100 generations	After 100 generations	After 100 generations
N-F	0.44	0.48	0.72	0.72
PD-D-T	0.64	0.64	0.88	1
D1-D2	0.2	0.2	0.2	0.2
T1&2-T3	0.48	0.48	1.12	1.2

TABLE 5 | Likelihood of optimality *Lopt(k)*.

Fault type	Traditional GA mutation rate = 0.01		Improved GA	
	After 100 generations	After 100 generations	After 100 generations	After 100 generations
N-F	0.2	0.2	1	1
PD-D-T	0	0	0.6	1
D1-D2	1	1	1	1
T1&2-T3	0	0	0.4	0.8
Average	0.25	0.3	0.75	0.95

200th generations for both the TGA and the IGA. **Table 5** shows the likelihood of optimality in the 100th and 200th generations for both TGA and the IGA. As can be seen from **Table 3**, compared with TGA, the average fitness values of IGA after 200 generations have increased 2.4, 1.7, 6.6% when detecting N-F, PD-D-T, and T1&2-T3, respectively. Besides, the average fitness values of IGA after 100 generations are higher than that of TGA after 200 generations, which proves that IGA has better search capability than the TGA. **Table 4**

TABLE 6 | Accuracy of different methods for the test samples.

		Accuracy	—	—	Accuracy
Traditional GA mutation rate = 0.01	Run 1	24/25	Improved GA	Run 1	25/25
	Run 2	24/25		Run 2	25/25
	Run 3	24/25		Run 3	24/25
	Run 4	23/25		Run 4	25/25
	Run 5	23/25		Run 5	25/25
Average Accuracy		94.4%	—	—	99.2%

TABLE 7 | Result comparison with other methods.

Method	IEC ratios (%)	Dual triangle (%)	SVM (%)	CVA (%)	Proposed method (%)
Accuracy	60	52	88	96	99.2

shows the average number of evolution leaps of the IGA is higher than the TGA, which indicates that the IGA has a continuous change in solution from one generation to the next. **Table 5** shows the probability of obtaining the optimal solution, the average likelihood of optimality of the IGA is 95%, compared to 30% of the TGA, which guarantees a feasible solution. The above results show that the IGA can get the optimal solution stably and reliably. For the test, **Table 6** shows the recognition accuracy of different methods for the test samples. The results in **Table 6** show that the proposed method has the best performance in transformer fault diagnosis with an average identification accuracy of 99.2%, compared to 94.4% of the method based on the TGA and XGBoost. Combined with the results of the test samples and the above results, it is shown that the proposed method can effectively and reliably identify transformer faults.

3.2 Comparison With Other Methods

In **Table 7**, the performance of the proposed method is also compared with other methods. The compared methods include DGA methods and intelligent transformer fault diagnosis methods, including IEC ratios, dual triangle, support vector machine, and common vector approach. The common vector approach has been recently introduced for the transformer fault diagnosis, and the support vector machine is a commonly used algorithm for transformer fault diagnosis. The result shows that the accuracy of DGA methods is relatively low, such as IEC ratios with an accuracy of 60% and the dual triangle method with an accuracy of 52%. In contrast to the DGA methods, intelligent transformer fault diagnosis methods based on intelligence algorithms and DGA methods have remarkable performance. The accuracy of the transformer fault identified by the proposed method is the highest, reaching 99.2%. Compared to 88% for support vector machines and 96% for CVA, the diagnostic accuracy of the proposed method was improved by 11.2 and 3.6%, respectively. The result proves that the proposed method can effectively improve the accuracy of transformer fault identification.

4 CONCLUSION

A novel and effective transformer fault diagnosis system based on the IGA-XGBoost is conducted to diagnose transformer fault types and verified in this paper. The modifications improve the global search capability of the IGA, and the IGA can get the optimal combined solution of input feature selection and the XGBoost classifier optimization reliably and stably. Based on the IGA, the IGA-XGBoost can accurately deal with different recognition problems, including N-F, PD-D-T, D1-D2, and T1&2-T3. Due to the excellent performance of the IGA-XGBoost, the average accuracy of the proposed transformer fault diagnosis system has been improved to 99.2%. Compared to IEC ratios, Dual triangle, SVM, and CVA, the simulation results demonstrate that the proposed method can be a reliable solution for transformer fault diagnosis.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: doi: 10.1016/j.epsr.2020.106346.

AUTHOR CONTRIBUTIONS

The individual contributions of authors are as follows: data curation, MZ; methodology, ZL; supervision, XC; validation, YH; writing (original draft), ZW. All authors have read and agreed to the published version of the manuscript.

FUNDING

This research is supported by Fujian Province Young and Middle-aged Teacher Education and Research Project (Grant No. JAT190592), China.

REFERENCES

Bhalla, D., Bansal, R. K., and Gupta, H. O. (2012). Function Analysis Based Rule Extraction from Artificial Neural Networks for Transformer Incipient Fault Diagnosis. *Int. J. Electr. Power Energ. Syst.* 43, 1196–1203. doi:10.1016/j.ijepes.2012.06.042

Castro, A. R. G., and Miranda, V. (2005). Knowledge Discovery in Neural Networks with Application to Transformer Failure Diagnosis. *IEEE Trans. Power Syst.* 20, 717–724. doi:10.1109/TPWRS.2005.846074

Chen, T., and Guestrin, C. (2016). XGBoost. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 785–794. doi:10.1145/2939672.2939785

Chen, Xi., Chen, W., and Gan, D. (2010).Properties and Gas Production Law of Surface Discharge in Transformer Oil-Paper Insulation. In 2010 Annual Report Conference on Electrical Insulation and Dielectic Phenomena. IEEE, 1–4. doi:10.1109/CEIDP.2010.5724049

Cheng, L., and Yu, T. (2018). Dissolved Gas Analysis Principle-Based Intelligent Approaches to Fault Diagnosis and Decision Making for Large Oil-Immersed Power Transformers: A Survey. *Energies* 11, 913. doi:10.3390/en11040913

Cheng, L., Yu, T., Wang, G., Yang, B., and Zhou, L. (2018). Hot Spot Temperature and Grey Target Theory-Based Dynamic Modelling for Reliability Assessment of Transformer Oil-Paper Insulation Systems: A Practical Case Study. *Energies* 11, 249. doi:10.3390/en11010249

Colorado, D., Hernández, J. A., Rivera, W., Martínez, H., and Juárez, D. (2011). Optimal Operation Conditions for a Single-Stage Heat Transformer by Means of an Artificial Neural Network Inverse. *Appl. Energ.* 88, 1281–1290. doi:10.1016/j.apenergy.2010.10.006

Daelemans, W., Hoste, V., De Meulder, F., and Naudts, B. (2003). Combined Optimization of Feature Selection and Algorithm Parameters in Machine Learning of Language. In European Conference on Machine Learning. Springer, 84–95. doi:10.1007/978-3-540-39857-8_10

Dhote, N., and Helonde, J. (2012). Diagnosis of Power Transformer Faults Based on Five Fuzzy Ratio Method. *WSEAS Trans. Power Syst.* 7, 12.

Dong, M., Yan, Z., and Taniguchi, Y. (2003).Fault Diagnosis of Power Transformer Based on Model-Diagnosis with Grey Relation. In Proceedings of the 7th International Conference on Properties and Applications of Dielectric Materials (Cat. No.03CH37417). IEEE, 1158–1161. doi:10.1109/ICPADM.2003.1218629

Fan, J., Wang, F., Sun, Q., Bin, F., Liang, F., and Xiao, X. (2017). Hybrid RVM-ANFIS Algorithm for Transformer Fault Diagnosis. *IET Generation, Transm. Distribution* 11, 3637–3643. doi:10.1049/iet-gtd.2017.0547

Fei, S.-w., Wang, M.-J., Miao, Y.-b., Tu, J., and Liu, C.-l. (2009). Particle Swarm Optimization-Based Support Vector Machine for Forecasting Dissolved Gases Content in Power Transformer Oil. *Energ. Convers. Manag.* 50, 1604–1609. doi:10.1016/j.enconman.2009.02.004

Fei, S.-w., and Zhang, X.-b. (2009). Fault Diagnosis of Power Transformer Based on Support Vector Machine with Genetic Algorithm. *Expert Syst. Appl.* 36, 11352–11357. doi:10.1016/j.eswa.2009.03.022

Fu Wan, W., Weigen Chen, W., Xiaojuan Peng, X., and Jing Shi, J. (2012).Study on the Gas Pressure Characteristics of Photoacoustic Spectroscopy Detection for Dissolved Gases in Transformer Oil. In 2012 International Conference on High Voltage Engineering and Application. IEEE, 286–289. doi:10.1109/ICHVE.2012.6357108

Ghoneim, S. S. M., Taha, I. B. M., and Elkalashy, N. I. (2016). Integrated ANN-Based Proactive Fault Diagnostic Scheme for Power Transformers Using Dissolved Gas Analysis. *IEEE Trans. Dielect. Electr. Insul.* 23, 1838–1845. doi:10.1109/TDEI.2016.005301

Jiang, X. Q., Gong, Y., Han, S., and Zhou, K. (2014). Application of the Improved Three-Ratio Method in Chromatographic Analysis of Locomotive Transformer Oil. *Amr* 1030-1032, 29–33. doi:10.4028/www.scientific.net/amr.1030-1032.29

Ke Meng, K., Zhao Yang Dong, Y., Dian Hui Wang, D. H., and Kit Po Wong, K. P. (2010). A Self-Adaptive RBF Neural Network Classifier for Transformer Fault Analysis. *IEEE Trans. Power Syst.* 25, 1350–1360. doi:10.1109/TPWRS.2010.2040491

Kirkbas, A., Demircali, A., Koroglu, S., and Kizilkaya, A. (2020). Fault Diagnosis of Oil-Immersed Power Transformers Using Common Vector Approach. *Electric Power Syst. Res.* 184, 106346. doi:10.1016/j.epsr.2020.106346

Li, J., Chen, X., and Wu, C. (2009).Application of Comprehensive Relational Grade Theory in Expert System of Transformer Fault Diagnosis. In 2009 International Workshop on Intelligent Systems and Applications. IEEE, 1–4. doi:10.1109/iwisa.2009.5072742

Lin, C. E., Ling, J.-M., and Huang, C.-L. (1993). An Expert System for Transformer Fault Diagnosis Using Dissolved Gas Analysis. *IEEE Trans. Power Deliv.* 8, 231–238. doi:10.1109/61.180341

Liu, Q., Huang, G., Mao, C., Shang, Y., and Wang, F. (2016).Recognition of Dissolved Gas in Transformer Oil by Ant colony Optimization Support Vector Machine. In 2016 IEEE International Conference on High Voltage Engineering and Application (ICHVE). IEEE, 1–4. doi:10.1109/ICHVE.2016.7800837

Liu, Z.-x., Song, B., Li, E.-w., Mao, Y., and Wang, G.-l. (2015). Study of "code Absence" in the IEC Three-Ratio Method of Dissolved Gas Analysis. *IEEE Electr. Insul. Mag.* 31, 6–12. doi:10.1109/MEI.2015.7303257

Mani, G., and Jerome, J. (2014). Intuitionistic Fuzzy Expert System Based Fault Diagnosis Using Dissolved Gas Analysis for Power Transformer. *J. Electr. Eng. Tech.* 9, 2058–2064. doi:10.5370/JEET.2014.9.6.2058

Miranda, V., and Castro, A. R. G. (2005). Improving the IEC Table for Transformer Failure Diagnosis with Knowledge Extraction from Neural Networks. *IEEE Trans. Power Deliv.* 20, 2509–2516. doi:10.1109/TPWRD.2005.855423

Mofizul Islam, S., Wu, T., and Ledwich, G. (2000). A Novel Fuzzy Logic Approach to Transformer Fault Diagnosis. *IEEE Trans. Dielect. Electr. Insul.* 7, 177–186. doi:10.1109/94.841806

Naresh, R., Sharma, V., and Vashisth, M. (2008). An Integrated Neural Fuzzy Approach for Fault Diagnosis of Transformers. *IEEE Trans. Power Deliv.* 23, 2017–2024. doi:10.1109/TPWRD.2008.2002652

Niu Wu, N., Xu Liangfa, L., and Hu Sanguo, S. (2010).Fault Diagnosis Method for Power Transformer Based on Ant colony -SVM Classifier. In 2010 The 2nd International Conference on Computer and Automation Engineering (ICCAE). IEEE, 629–631. doi:10.1109/ICCAE.2010.5451326

Nogami, T., Yokoi, Y., Ichiba, H., and Atsumi, Y. (1995). Gas Discrimination Method for Detecting Transformer Faults by Neural Network. *Elect. Eng. Jpn.* 115, 93–103. doi:10.1002/eej.4391150109

Saha, T. K., and Purkait, P. (2004). Investigation of an Expert System for the Condition Assessment of Transformer Insulation Based on Dielectric Response Measurements. *IEEE Trans. Power Deliv.* 19, 1127–1134. doi:10.1109/TPWRD.2004.829100

Shang, H., Xu, J., Zheng, Z., Qi, B., and Zhang, L. (2019). A Novel Fault Diagnosis Method for Power Transformer Based on Dissolved Gas Analysis Using Hypersphere Multiclass Support Vector Machine and Improved D-S Evidence Theory. *Energies* 12, 4017. doi:10.3390/en12204017

Singh, S., and Bandyopadhyay, M. (2010). Dissolved Gas Analysis Technique for Incipient Fault Diagnosis in Power Transformers: A Bibliographic Survey. *IEEE Electr. Insul. Mag.* 26, 41–46. doi:10.1109/MEI.2010.5599978

Souahlia, S., Bacha, K., and Chaari, A. (2012). MLP Neural Network-Based Decision for Power Transformers Fault Diagnosis Using an Improved Combination of rogers and Doernenburg Ratios DGA. *Int. J. Electr. Power Energ. Syst.* 43, 1346–1353. doi:10.1016/j.ijepes.2012.05.067

Sugihara, K. (1997).Measures for Performance Evaluation of Genetic Algorithms. In Proc. 3rd. joint Conference on Information Sciences. State College, Commonwealth of Pennsylvania: Citeseer, 172–175.

Tightiz, L., Nasab, M. A., Yang, H., and Addeh, A. (2020). An Intelligent System Based on Optimized Anfis and Association Rules for Power Transformer Fault Diagnosis. *ISA Trans.* 103, 63–74. doi:10.1016/j.isatra.2020.03.022

Wang, Z., Liu, Y., and Griffin, P. J. (2000). A Combined Ann and Expert System Tool for Transformer Fault Diagnosis. In *2000 IEEE Power Engineering Society Winter Meeting. Conf. Proc.* 2, 1261–1269. (Cat. No. 00CH37077) (IEEE).

Weigen Chen, W., Chong Pan, C., Yuxin Yun, Y., and Yilu Liu, Y. (2009). Wavelet Networks in Power Transformers Diagnosis Using Dissolved Gas Analysis. *IEEE Trans. Power Deliv.* 24, 187–194. doi:10.1109/TPWRD.2008.2002974

Xi, L., Wu, J., Xu, Y., and Sun, H. (2020). Automatic Generation Control Based on Multiple Neural Networks with Actor-Critic Strategy. *IEEE Trans. Neural Netw. Learn. Syst.* PP, 2483–2493. doi:10.1109/TNNLS.2020.3006080

Xu, W., Zhou, Z., Chen, H., and Wang, D. (1997). Fault Diagnosis of Power Transformers: Application of Fuzzy Set Theory, Expert Systems and Artificial Neural Networks. *IEE Proc. - Sci. Meas. Tech.* 144, 39–44. doi:10.1049/ip-smt:19970856

Yadaiah, N., and Ravi, N. (2011). Internal Fault Detection Techniques for Power Transformers. *Appl. Soft Comput.* 11, 5259–5269. doi:10.1016/j.asoc.2011.05.034

Yann-Chang Huang, Y-C., Hong-Tzer Yang, H-T., and Ching-Lien Huang, C-L. (1997). Developing a New Transformer Fault Diagnosis System through Evolutionary Fuzzy Logic. *IEEE Trans. Power Deliv.* 12, 761–767. doi:10.1109/61.584363

Yi, J.-H., Wang, J., and Wang, G.-G. (2016). Improved Probabilistic Neural Networks with Self-Adaptive Strategies for Transformer Fault Diagnosis Problem. *Adv. Mech. Eng.* 8, 168781401562483. doi:10.1177/1687814015624832

Yin, J., Zhu, Y., and Yu, G. (2011).Power Transformer Fault Diagnosis Based on Support Vector Machine with Cross Validation and Genetic Algorithm. In 2011 International Conference on Advanced Power System Automation and Protection. IEEE, 309–313. doi:10.1109/APAP.2011.6180419

Yuan, F., Guo, J., Xiao, Z., Zeng, B., Zhu, W., and Huang, S. (2019). A Transformer Fault Diagnosis Model Based on Chemical Reaction Optimization and Twin Support Vector Machine. *Energies* 12, 960. doi:10.3390/en12050960

Žarković, M., and Stojković, Z. (2017). Analysis of Artificial Intelligence Expert Systems for Power Transformer Condition Monitoring and Diagnostics. *Electric Power Syst. Res.* 149, 125–136. doi:10.1016/j.epsr.2017.04.025

Zeng, W., Yang, Y., Gan, C., Li, H., and Liu, G. (2011).Study on Intelligent Development of Power Transformer On-Line Monitoring Based on the Data of DGA. In 2011 Asia-Pacific Power and Energy Engineering Conference. IEEE, 1–4. doi:10.1109/appeec.2011.5749107

Zhang, L., and Zhan, C. (2017).Machine Learning in Rock Facies Classification: An Application of XGBoost. In International Geophysical Conference, Qingdao, China, 17-20 April 2017. Tulsa: Society of Exploration Geophysicists and Chinese Petroleum Society, 1371–1374. doi:10.1190/IGC2017-351

Zhang, X., Xu, Z., Yu, T., Yang, B., and Wang, H. (2020). Optimal Mileage Based Agc Dispatch of a Genco. *IEEE Trans. Power Syst.* 35, 2516–2526. doi:10.1109/tpwrs.2020.2966509

Zhu, X., Xiong, J., and Liang, Q. (2018). Fault Diagnosis of Rotation Machinery Based on Support Vector Machine Optimized by Quantum Genetic Algorithm. *IEEE Access* 6, 33583–33588. doi:10.1109/ACCESS.2018.2789933

Conflict of Interest: The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Energy Research

ORIGINAL RESEARCH
published: 13 October 2021
doi: 10.3389/fenrg.2021.730058

An Association Rules-Based Method for Outliers Cleaning of Measurement Data in the Distribution Network

Hua Kuang[1], Risheng Qin[2], Mi He[3], Xin He[2], Ruimin Duan[2], Cheng Guo[2] and Xian Meng[2]*

[1]Yunnan Power Grid Co., Ltd., Kunming, China, [2]Electric Power Research Institute of Yunnan Power Grid Company Ltd., Kunming, China, [3]Kunming Power Supply Company of Yunnan Power Grid Co., Ltd., Kunming, China

For any power system, the reliability of measurement data is essential in operation, management and also in planning. However, it is inevitable that the measurement data are prone to outliers, which may impact the results of data-based applications. In order to improve the data quality, the outliers cleaning method for measurement data in the distribution network is studied in this paper. The method is based on a set of association rules (AR) that are automatically generated form historical measurement data. First, the association rules are mining in conjunction with the density-based spatial clustering of application with noise (DBSCAN), k-means and Apriori technique to detect outliers. Then, for the outliers repairing process after outliers detection, the proposed method uses a distance-based model to calculate the repairing cost of outliers, which describes the similarity between outlier and normal data. Besides, the Mahalanobis distance is employed in the repairing cost function to reduce the errors, which could implement precise outliers cleaning of measurement data in the distribution network. The test results for the simulated datasets with artificial errors verify that the superiority of the proposed outliers cleaning method for outliers detection and repairing.

Keywords: association rules, outliers cleaning, outliers detection, outliers repairing, measurement data, distribution network

OPEN ACCESS

Edited by:
Qiuye Sun,
Northeastern University, China

Reviewed by:
Zhijun Qin,
Guangxi University, China
Shunbo Lei,
University of Michigan, United States
Xuguang Hu,
Northeastern University, China

****Correspondence:***
Risheng Qin
qinrisheng2020@126.com

Specialty section:
This article was submitted to Smart Grids, a section of the journal Frontiers in Energy Research

Received: *24 June 2021*
Accepted: *27 September 2021*
Published: *13 October 2021*

Citation:
Kuang H, Qin R, He M, He X, Duan R, Guo C and Meng X (2021) An Association Rules-Based Method for Outliers Cleaning of Measurement Data in the Distribution Network. Front. Energy Res. 9:730058. doi: 10.3389/fenrg.2021.730058

INTRODUCTION

With the evolution of smart grids, the intelligent monitoring equipment and system are becoming an integral component of the distribution network, collecting a substantial volume of data in order to manage the status and provide timely updates in the network (Alimardani et al., 2015; Wang et al., 2018). Among them, the supervisory control and data acquisition (SCADA) system provides a large number of operation data and analysis results, which brings great convenience for operators to evaluate the planning and operation of distribution system. For instance, the data structure is complex, many types of the data, and the sampling period/frequency of data are also different. For distribution network dispatching control system, poor quality data may lead to wrong decisions, which will have a great impact on the stable operation of power grid. Hence, it is essential that to clean the outliers of measurement data in the distribution network.

The distribution network is an important part of production, transmission, and consumption, which plays a critical role in the delivery of electric power. In the planning and operation of distribution network, the availability of accurate measurement data has a considerable impact on dispatching operations and control of the distribution network. For instance, the analysis of measurement data in the distribution network can assist in taking action against fault detection,

dispatching, load forecasting, power quality, tariff settings, and so forth. (Hayes et al., 2018; Cai et al., 2021; Wang et al., 2019). Moreover, it solves the problems that distribution networks frequently face in terms of integrated energy planning, distributed energy storage, and demand-side management, respectively (Thams et al., 2018; Liu et al., 2019). Generally, the majority of the researches in the distribution network, for the analysis and prediction of the measurement data, is focus on the feature selection or parametric optimization of the model (Liu et al., 2020). However, due to the complex topology features and communication disturbances, the accuracy of distribution network measurement data is not always satisfactory, making it susceptible to data anomalies such as outliers or missing data (Shi et al., 2019). To fill in missing data, denoise while detecting outliers, and repair inconsistencies, data cleaning is the first and most crucial step. Obviously, it has a decisive influence on the final result: if the dataset is incomplete in terms of data cleaning and preprocessing. This means that the established analysis and prediction model will not be accurate and efficient, which may no longer be suitable for the planning and operation of distribution system. For instance, due to external disturbances, data recorded in smart electric meters is abruptly modified because a transmission error for control commands, such as electric quantity or associated parametric information is reset to outliers, or even data missing (Nascimento et al., 2012). And in DC microgrids, the large-scale converters with inhomogeneous initial values are widely appeared due to soft-starting operation, which make the input-output maps error will be large (Wang et al., 2021). Under these circumstances, efficient preprocessing via data cleaning aids in improving the quality and accuracy of subsequent analysis and decision-making outcomes, which can successfully guide the planning and operation of the distribution network.

Researchers have extensively conducted many outliers cleaning studies to improve the data quality and decision-making results, including outliers detection and repairing. For outliers detection, with the rapid development of machine learning technology, many machine learning algorithms have been utilized to improve the accuracy in power systems. In literature (Nemati et al., 2018), a constraint and association rule-based current transmission capability forecasting method was proposed for outliers detection in substation metering equipment. However, this model is complex and computationally intensive, which is not suitable for the detection of bad data in a large number of transformer districts. In literature (Esmalifalak et al., 2014), support vector machine (SVM) has been investigated for detecting the outliers injected into the measurement data from power grid. Since SVM is a supervised learning method, it necessitates labeling the data in order to train the model. However, in practice, obtaining a considerable volume of tagged data is difficult. In literature (Thang et al., 2011), a density-based DBSCAN algorithm was used for detecting the network traffic outliers of electricity meters, which dataset may include multiple traffic types with different characteristics. It has a high level of outliers detection performance, but there are difficulties in finding its parameters (epsilon and minpts) when the multidimensional feature data is taken into account. In literature (Li et al., 2018), the isolation forest (IF) algorithm was proposed to detect the outliers, and the backpropagation neural network (BPNN) algorithm was used for predicting and repairing the outliers. However, IF algorithm is usually suitable for detecting global outliers, but not for detecting local outliers.

Traditionally, researchers have concentrated more on a basic and easy to repair statistical estimate method for outliers repairing. Still, mining a deep relationship between data is difficult, and the repairing results are not ideal (Waal et al., 2001). By contrast, machine learning (also includes deep learning) methods is a very effective technology, which could easily recognize the outliers through the linear or nonlinear pattern relationships and the repairing results could more accurately. For instance, in literature (Qu et al., 2016), a hierarchical clustering algorithm based on the clustering using representatives (CURE) was proposed for the outliers detection the repairing, which could confirm the normal value boundary samples from historical data. However, when the volume of data is large, the time complexity is poor and precision is low with the hierarchical clustering algorithm, which make a challenge to determine the ideal boundary sample number. In literature (Hu et al., 2021) a data recovery method based on generative adversarial networks (GANs) was proposed for safe and efficient operation in the pipeline network, which could accurately recover incomplete pressure data caused by the device or communication aspect. But there are still some difficulties when the complete data pairs is no provided in the training process. On the other hand, to produce good repairing results, a metric learning and a cost functional model are proposed to estimate data repairing efficiency while taking sample distances into account (Li et al., 2019). Distance is a term that describes the dissimilarity of two input samples. Among them, the most frequently used technique is the Euclidean distance. However, the Euclidean distance takes neither the correlation of the features nor the different weights of features into account, which may not reflect the real nature of the problem, and distorts the true dissimilarity between samples. To address this issue, in literature (Maesschalck et al., 2000), the Mahalanobis distance idea was defined to use the similarity metric as a substitution to perform better. In another example (Yan et al., 2020), the adoption of the Mahalanobis distance improves the classic k-nearest neighbor (KNN) outliers identification method, resulting in increased accuracy and a lower false detection rate. However, the repair of outliers has not been considered in this model. Furthermore, most of the models stated above focus on specific application scenarios and do not process real-time data from the distribution network system. Moreover, most of the outliers cleaning methods aforementioned presumed the underlying population distribution before the step of data cleaning. However, in real-word data, a hypothesis about an underlying population is a statement that may be true or false.

In power grids, a huge amount of historical measurement data from various distribution stations is available, which could provide valuable information for detecting and repairing outliers. Furthermore, the association rules learning is a popular and well data mining method for discovering relations

between variable features. Our motivation is to investigate how to capitalize on the historical data for outliers cleaning, including outliers detection and repairing, to achieve expected performance. An association rules-based method for outliers cleaning is given in this work to mine the information which whereas the assumptions of underlying population about the data is not required. In outliers detection, we adopt the density-based spatial clustering of application with noise (DBSCAN), k-means and Apriori technique to generate the association rules. After the outliers detection, the distance-based model is designed with Mahalanobis distance to repair to outliers. Various tests are carried out on data sets with simulated errors to evaluate the good performance of the proposed method. The test results indicate that the proposed method can effectively identify outliers in the distribution network's measurement data while achieving accurate data repairing. The proposed method detects the outliers with a F1-Score (a metric combine precision and recall) of 96%, even in the condition with a high anomaly rate. The F1-Score indicates how well precision and memory are balanced. Furthermore, the correlation between features of measurement data is also computed to detect and repair the outliers, thus improving the method's accuracy. The main contributions of this work could be summarized as follows two aspects:

1) This paper introduces an outlier detection and repairing technique based on association rule. The proposed technique uses the information provided from historical measurement data, whereas the assumptions of underlying distribution about the measurement data is not required.
2) The distance-based model is adopt for outliers repairing, which describes the similarity between outlier and normal data by the Mahalanobis distance. It estimates the outliers according to the normal data within historical data, which is employed to improve the estimation accuracy.

The remainder of the paper is organized in the following manner. *Problem Statement* discusses the measurement problems in the distribution network and data anomalies. The proposed methodology has been presented in *Preliminaries* and *The Proposed Model for Outliers Detection and Repairing*. Simulation results are provided in *Experiment and Analysis*, and the concluding remarks are summarized in *Conclusion*.

PROBLEM STATEMENT

Measurement Problems in Distribution Network

Through SCADA system, a large amount of operating data is continuously collected, uploaded, and formed into big data for the distribution network, which provides abundant data resources for big data analysis (Ye et al., 2010; Song et al., 2013). And the data collected by SCADA has the following characteristics: large amount, high dimensions, and complex data types. Then the most common problems encountered in measurement data are the absence of data (nulls values and zeros), change in level and spikes (points more than N times the standard deviations away from the series mean), and generally are called outliers. Therefore, in order to improve the accuracy of the analysis and decision-making results based on measurement data, how to clean and repair outliers form the measurement data in distribution network is a challenge faced by distribution system.

The Source of Abnormal Data in Distribution Network

The process of collecting measurement data from the distribution network involves many components such as metering equipment, metering centers, and communication systems. However, if a malfunction occurs in any measurement channel, it can lead to data anomalies (Chen et al., 2010). For example, the failure of smart electricity meters, noise interference, data transmission errors, and abnormal power consumption will cause these collected data to become outliers or data missing. Generally, there are three potential sources of data anomalies in distribution network measurement data:

1) Metering equipment. The measuring equipment from abnormal operating conditions may lead to errors in measurement (Yan et al., 2015; Chen et al., 2010). In particular, the magnetic bias phenomenon in potential transformers (PT) and current transformers (CT) equipment would cause measurement errors (Mccamish et al., 2016). Also, the non-synchronous problem on data collection could cause errors since the sampling time of some devices is asynchronous (Liu et al., 2020). In particular, all forms of metering and communication equipment are constantly exposed to unknown conditions. They are vulnerable to the effects of real-world circumstances, which typically have a high failure rate. Meanwhile, the operation in the monitoring and communication equipment can not be carried out smoothly when a fault occurs. In that situation, erroneous or missing data will be recorded.
2) Distribution network. Control operations and faults in the distribution network have a significant impact on the accuracy of measurement data. Temporary inrush current interference caused by switchgear such as circuit breakers may cause temporary outliers to appear in some measurements when adjusting the operation of the distribution network. In any fault event, the metering equipment may fail to function properly, resulting in measurement issues.
3) Communication systems. Due to the distribution network's complex topology and geographical environment, local communication links usually use low-power and lossy networks in power distribution networks. This type of network is prone to data packet loss. Also, the reliability of distribution network data transmission is affected by the communication links. The way of communication will also affect the reliability of data transmission in the distribution network. Due to cost constraints, most distribution topologies use communication methods such as distribution carrier waves, Zigbee wireless technology, and industrial wiring

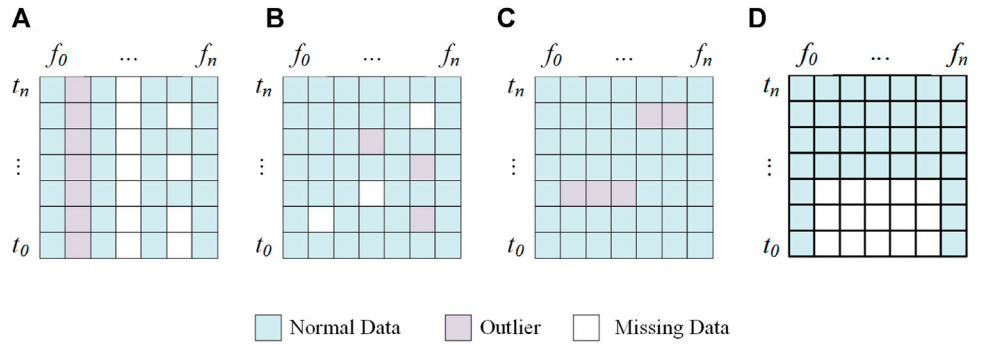

FIGURE 1 | The distribution of abnormal voltage data for several reasons: **(A)** Abnormal voltage data cause by Metering equipment; **(B)** Abnormal voltage data cause by faults; **(C)** Abnormal voltage data cause by operation control; **(D)** Abnormal voltage data caused by the communication system.

(Pei et al., 2010). These communication methods are less reliable and often break codes when the channel is exposed to heavy electromagnetic interference, resulting in missing data.

All of the above issues may produce anomalous data, causing the quality of the data to be inconsistent and reduce usability. Therefore, it is necessary to clean the data before using it for analysis and utilization. The prominent data anomalies in the existing distribution network data are missing data and outliers. The term "data missing" applies when the collected value is null or contains an invalid value. In contrast, outliers occur when the collected value deviates from normal data. The value exceeds the acceptable range of change (data is too big or too small) and maintains a certain time pattern without repetition.

Figure 1 illustrates the distribution of abnormal voltage data in the distribution network for various purposes. **Figure 1A** shows the abnormality caused by the malfunction of the metering equipment. The characteristic feature of this phenomenon is that some observations are outliers or missing values, which do not last for a long time but occur frequently. **Figure 1B** shows the data abnormality caused by the failure of a terminal monitoring point. It is characterized by continuous data anomalies or single-point data anomalies occurring at a single point of observation. **Figure 1C** shows the data anomalies caused by excited inrush disturbances and automation equipment actions at controller monitoring points. This abnormality is characterized by short-term outliers in some observations, i.e., retained for a very small period of time. **Figure 1D** shows data anomalies caused by faults in the communication system of the sub-stations. It is defined by a partial loss of temporal data at various intervals and is typically retained for a short period of time.

Outliers Cleaning in Distribution Network

The above issues might pollute the measurement data, which make it not suitable used directly for distribution system planning and operation. Therefore, the data preprocessing is an essential process before using data for analysis and decision-making, and the outliers detection is the most important part of the process. Generally speaking, a good outlier detection algorithm should be able to identify outliers correctly, and would no have any response to the normal data. As shown in **Figure 2**, a dataset of the voltage amplitude, which received from sensors and contains four outliers and one missing data, and highlight it in the figure. The aim of the outliers detection is to find the highlight points and mark it with lables, which is the kernel of the data cleaning.

Association rules learning are a rule-based machine learning method, which is a research focus of the data mining and analysis. In an method for automatically generating association rules, it mainly includes three important steps: data denoising, data discretization and rule mining. Furthermore, how to select the sub-algorithm is a critical step. In this paper, the density-based algorithm, DBSCAN, is chosen in the data denoising step, which has a excellent result in denoising and high scalability. Then in the data discretization step, the distance-based algorithm, K-means is used since its precise classification result and high computation efficiency. And in the rule mining step, Apriori algorithm is selected because of its high stability and flexible extension ability. With that in mind, we presents an outliers cleaning method based on association rules, which could found the implicit relationship between features from the historical measurement data and pick up the valuable information on outliers detection and repairing. For the outliers detection, the DBSCAN, K-means and Apriori algorithm are chosen for generating the association rules from historical data, which make the detector more flexible and accurate. For the outliers repairing, the repairing cost is chosen with a distance-based model. And the Mahalanobis distance is chosen to use for constructing a data repairing cost function, which could reduce the errors.

PRELIMINARIES

DBSCAN Clustering Algorithm

DBSCAN is an unsupervised machine learning clustering algorithm that could be used for data classification with a nonlinear density structure (Chen et al., 2021; Chipade et al., 2021). The algorithm treats the data as points in space and clusters them based on density magnitude, allowing clusters of arbitrary shapes to be found in a noisy space. The basic idea of

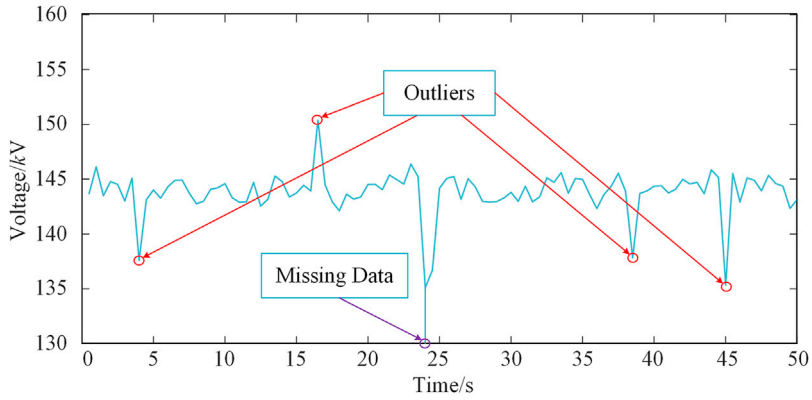

FIGURE 2 | The data anomaly plot.

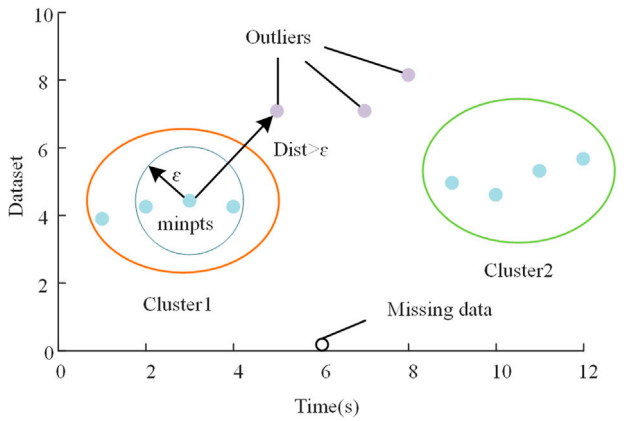

FIGURE 3 | The process of DBSCAN clustering.

DBSCAN is to introduce neighborhood and density connectivity concepts, explore the data points, and use density connectivity to grow clusters until outliers split them. The DBSCAN clustering algorithm can be adapted to any form of clustering. It can filter out noisy outliers in space, making it ideal for outliers detection in the distribution network.

Assume a historical measurements datasets $\boldsymbol{D} = \{x_1, x_2, \ldots, x_n\}$ be the numerical attributes of observations with rows $i \in [1, 2, \ldots, n]$. For any observation $x_i \in \boldsymbol{D}$ has a timestamp. And it contains m features, as given by **Equation 1**.

$$\boldsymbol{x_i} = \{\boldsymbol{f_{i1}}, \boldsymbol{f_{i2}}, \ldots, \boldsymbol{f_{im}}\} \tag{1}$$

where f_{ij} represents the jth feature of the ith data.

Meanwhile, let ε be the neighborhood distance parameter of DBSCAN. For each point $x_i \in \boldsymbol{D}$, the ε - neighborhood set is defined using **Eq. 2**.

$$\boldsymbol{N_\varepsilon}(\boldsymbol{x_i}) = \{\boldsymbol{x_j}\, \varepsilon\, \boldsymbol{D} | \boldsymbol{dist}(\boldsymbol{x_i}, \boldsymbol{x_j}) \leq \varepsilon\} \tag{2}$$

Then, for any point $x_i \in \boldsymbol{D}$, the core point should be satisfied *via* **Eq. 3**.

$$|\boldsymbol{N_\varepsilon}(\boldsymbol{x_i})| \geq \boldsymbol{minpts} \tag{3}$$

where, $|N_\varepsilon(x_i)|$ is the count of elements in the $N_\varepsilon(x_i)$. And *minpts* is the minimum number of points in ε -neighborhood.

If a point $x_i \in N_\varepsilon(x_i)$ satisfy **Eq. (2)**, x_i is directly-density reachable from the point x_j. As shown in **Figure 3**, points that are outside the range of clustering are considered outliers. For clarity, **Algorithm 1** explains the step-by-step procedure of the DBSACN clustering algorithm.

Algorithm 1. : Density-based spatial clustering of applications with noise(DBSCAN).

Density-based spatial clustering of applications with noise(DBSCAN)

Input: datasets D, ε-neighborhood, and minpts.
Output: the density-based clusters $\mathbf{C}_k$ and the label set **L**.
Feed the dataset into **D** and all the points are label as unvisited
Initialize the set of core points: $H = \emptyset$, the number of clusters: $k = 0$, and the points which is unvisited: $P = \mathbf{D}$
for i =1 to n **do**
  $N_\varepsilon(x_i) = \text{find}(dist \leq \varepsilon)$, $|N_\varepsilon(x_i)| = \text{count}(N_\varepsilon(x_i))$
  if $|N_\varepsilon(x_i)| \geq minpts$ **then**
    Feed x_i into the set of core points: $H = H \cup \{x_i\}$
  end if
end for
while $H \neq \emptyset$ **do**
  Update the set of points which are unvisited: $P_{old} = P$
  Select a core point randomly: $cp \in H$, and initialize the queue of core points: $Q = \langle cp \rangle$
  $P = P \setminus cp$
  while $Q \neq \emptyset$ **do**
    Pick out the first core point q from Q
    if $|N_\varepsilon(x_i)| \geq minpts$ **then**
      $\triangle = |N_\varepsilon(q)| \cap P$; Feed $\triangle$ into Q, $P = P \setminus \triangle$
    end if
  end while
  $k = k + 1$, $C_k = P_{old} \setminus P$, $H = H \setminus C_k$
  return result
end while

Association Rules Mining

Association rule learning is a rule-based machine learning method used to mine frequent patterns, correlations, or causal structures between itemsets. It is intended to determine valuable rules based on the frequency of occurrence between itemsets in a database (Rauch, 2005; Chengyu et al., 2016). In data cleaning, the association rules algorithm is applied for mining the relationship between various features of the measurement data in the power system.

At first, the obtained features should be discretized to improve the robustness of association rules to outliers. After data discretization, a datasets **D** with size $n \times m$, which is a 2-dimensional real-valued matrix, is converted to a Boolean

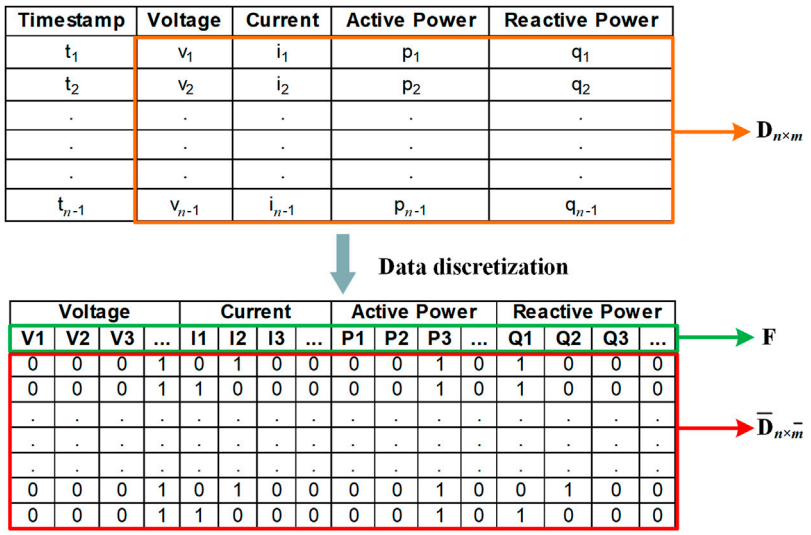

FIGURE 4 | The process of data discretization.

matrix $\overline{D}$ with size $n \times \overline{m}$. When a feature is in the specified interval, the value in the Boolean matrix is labelled as 1. If not, it is labelled as 0. **Figure 4** illustrates a portion of the datasets before and after data discretization.

Let $F = \{F_1, F_2, \ldots, F_n\}$ be the itemsets of $\overline{D}$. Each observation $\overline{x_i} \in \overline{D}$ may contain one or more items. The aim is to search for the most frequent patterns of items from the datasets for generating association rules. With this in mind, an association rule between F_A and F_B, can be defined as **Eq. 4**.

$$F_A \Rightarrow F_B \tag{4}$$

where F_A, F_B are itemsets, and $F_A \subset F$, $F_B \subset F$.

Since the magnitude of support and confidence value is commonly used to assess the effectiveness of an association rule. The following **Eq. 5** represents a definition to support an association rule between F_A and F_B:

$$Sup(F_A \Rightarrow F_B) = \frac{coun(F_A \cup F_B)}{|\overline{D}|} \tag{5}$$

where *count* is the number of occurrences of an item in data $\overline{D}$; $F_A \cup F_B$ is the coexistence of F_A and F_B; and $|\overline{D}|$ is the total count of itemsets.

The confidence value indicates the reliability of the association rule. The confidence of an association rule between and can be defined as follows:

$$Con(F_A \Rightarrow F_B) = \frac{coun(F_A \Rightarrow F_B)}{Sup(F_A)} \tag{6}$$

In this work, we use Apriori algorithm is employed to search for the itemsets frequency in the complete transaction set. In this approach, the ones with more than the minimum support and the minimum confidence are used as strong association rules. These rules give high confidence and strong support greater than or equal to a user-specified minimum confidence threshold and a minimum support threshold. The process of the algorithm is explained in **Algorithm 2**.

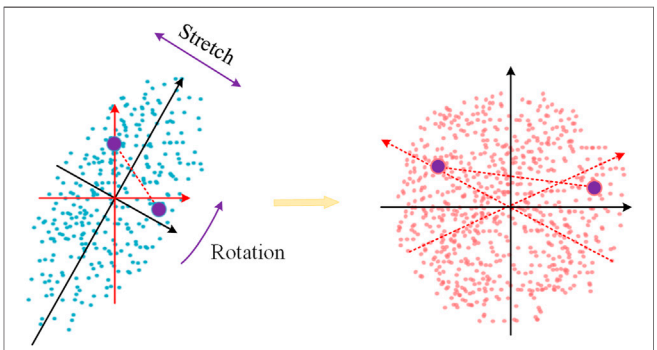

FIGURE 5 | An illustration of Mahalanobis distance.

Algorithm 2. : Association Rule Mining with Apriori.

Association Rule Mining with Apriori

Input: the discrete matrix $\overline{D}$, the itemsets F, minSup, and minCon.
Output: the frequent itemsets FI.
1: Feed the discrete matrix into $\overline{D}$ and feed the itemsets into F
2: Initialize the length of frequent itemsets: k = 1, initialize the candidate itemsets of size k: CI_k, and initialize the frequent itemsets of size k: FI_k
3: L1 = find(large frequent 1-itemsets)
4: **for** k = 1; $L_1 \neq \emptyset$; k++ **do**
5: $CI_{k+1} = GenerateCandidates\ (FI_k)$
6: **for** each observation $\overline{x_i} \in \overline{D}$ **do**
7: Increment count of candidates in CI_{k+1} that are contained in x_i
8: **end for**
9: $FI_{k+1} = candidates\ in\ CI_{k+1}\ with\ Sup \geq minSup\ and\ Con \geq minCon$
10: **end for**
11: **return** FI = $\cup_k FI_k$

Mahalanobis Distance

The Euclidean distance is the most common metric of distance in data science, which describes the straight-line distance between two points in Euclidean space. Consider the case where two or more variables are linked. The axes are no longer at right angles in this

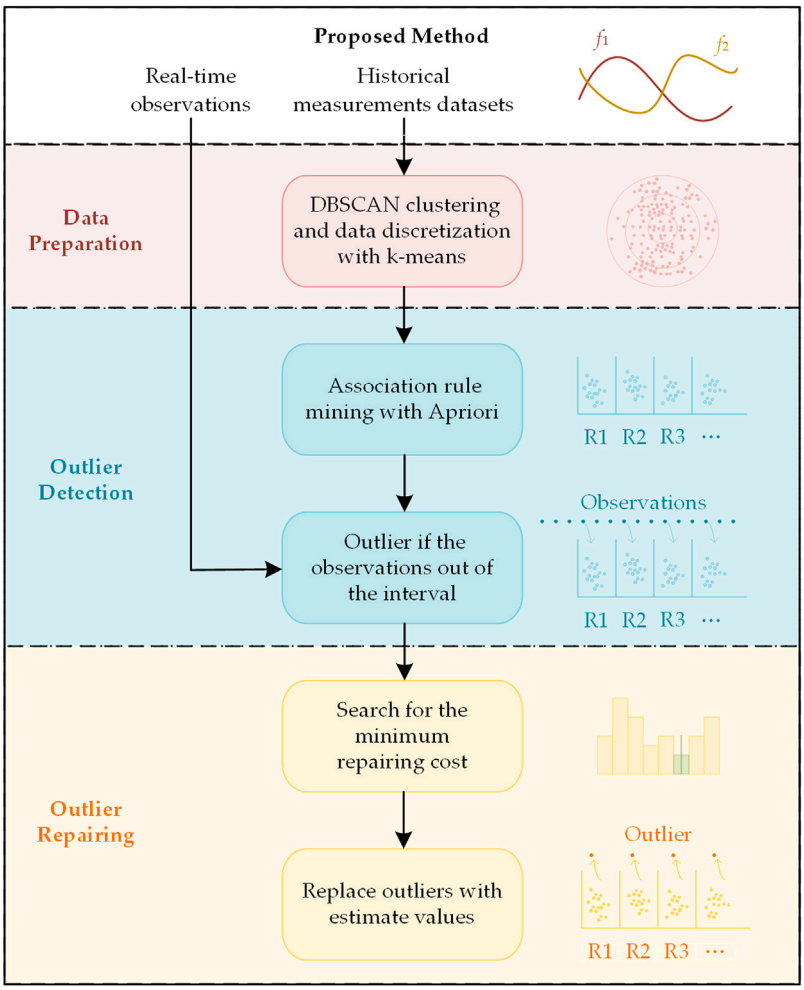

FIGURE 6 | The schema of the proposed method.

situation, and measurements are no longer possible. The Euclidean distance cannot represent the real distance between feature vectors. To mitigate this issue, Mahalanobis distance is implemented. The Mahalanobis distance measures the distance between two points in multivariate space (Maesschalck et al., 2000). It calculates distances between points, including correlated points for multiple variables.

For a given sample set, in order to integrate the correlation between the sample points, the distance between numerical data features can be calculated using the Mahalanobis distance metric to measure the similarity between samples (Liu, et al., 2020). For a feature, the variation between two observations can be specified by **Eq. 7**.

$$\begin{aligned}\mathbf{M}_d(x_1, x_2) &= \sqrt{(x_1 - x_2)^T \mathbf{S}^{-1} (x_1 - x_2)} \\ &= \sqrt{(x_1 - x_2)^T \mathbf{M} (x_1 - x_2)}\end{aligned} \tag{7}$$

where S is the covariance matrix, and $M = S^{-1}$, if the two samples have similar or identical characteristics, the martingale distance should be small or even zero.

Figure 5 illustrates the transformation in the two-dimension space. Here, the blue and pink dots represent the original and transformed data. For presented data points, the correlation of the two features causes the oval shape of the original distribution. If we apply the Euclidean distance, it will not reflect the real dissimilarity of the data. While calculating Mahalanobis distance, the eclipse is first transformed to a standardized circle with a radius equal to 1, and then the Euclidean distance in the transformed space is calculated. Meanwhile, computing the distance, the influence of correlation is offset by the transformation.

THE PROPOSED MODEL FOR OUTLIERS DETECTION AND REPAIRING

Since historical measurement data from SCADAS is required processing, using this information, the proposed model generates a list of association rules to evaluate the correlation between

FIGURE 7 | The RBS with association rules.

distinct variables. The overall flowchart of the proposed method is shown in **Figure 6**. The proposed model consists of three stages: data preparation, outliers detection and outliers repairing.

1) In the data preparation process, it has a task with DBSACN clustering algorithm to eliminate ineffective information, which purposes is to find the obvious abnormal data like null or missing values and so on. Then prepared dataset (**D**) will be discretized with k-means clustering algorithm, which purposes is to get the discrete dataset ($\overline{D}$) to generate association rules.
2) In the outliers detection process, it is necessary to mine the association rules in historical data for detecting outliers. The association rules will be mined from the dataset ($\overline{D}$) with Apriori algorithm. Correspondingly, the intervals of the features will be define by the association rules, which the features are distributed. The newly obtained observation will be compared with the intervals which defined from the association rules. If these real-time observations out of the intervals (i.e., identified by the association rules), they will be flagged as outliers. In this paper, we use 1 as the label for outlier, and 0 as the label for normal data.
3) In the outliers repairing process, all the sample points would be mapped into a feature space, which determining by the association rules. Subsequently, a novel cost function is constructed and used for data repairing, and the outlier will be repaired with the value which have the minimum repair cost. In this paper, the distance metric is formed with Mahalanobis distance, and similar normal data related to the query outliers are retrieved.

Data Preparation

In practice, there are some anomalies in the historical measurement data from the distribution network, which may bring invalid/incorrect information. Therefore, it is necessary to process the historical data before mining the association rules. We employ the DBSCAN clustering algorithm as the noise detector and design a procedure to dispose of the anomalies in the historical measurement data. The DBSCAN algorithm divides these observations into several clusters and outliers. The parameter of and *minpts* is chosen based on the silhouette coefficient. Next, the data points in each cluster are labelled as 0, while the outliers are labelled as 1. Furthermore, to generate the association rules, we use k-means for data discretization.

Outliers Detection

The results from the association rules are used for detecting the outliers. In any case, the outliers detection of each feature is based on comparisons between the new real-time observations and the association rules generated from all historical measurement data. In the final analysis, if a real-time measurement mismatches the interval defined by the associated rule, it will be flagged as an outlier.

For example, assume association rules are generated as follows:

$$\{F_A = [F_{A1\,min}, F_{A1\,max}) \Rightarrow F_B = [F_{B1\,min}, F_{B1\,max})\}$$

For a new observation O_t which contains the same features $(F_A \text{ and } F_B)$:

$$\text{If } O_t(F_A) \in [F_{A1\,min}, F_{A1\,max}), \text{while } O_t(F_B) \notin [F_{B1\,min}, F_{B1\,max})$$

According to the association rule, the new observation is compared to previous observations that have the same features. The O_t stays out of the intervals, which signify that O_t is an outlier, and the current observation would be labeled as 1.

Outliers Repairing

As shown in **Figure 7**, after outliers detection, for any point $x_i \in D$, it would be allocated in a feature space divided by the association rules, which we called "data binding". When an observation of one feature is marked as an outlier, it is necessary to calculate the estimated value of the outlier. For an abnormal observation in the "rule box space (RBS)", the point with the highest similarity to its attribute should fall in the same sub-RBS.

For example, assume the presence of outliers for the voltage magnitude, and the following Rule1 are generated with the highest confidence:

$$\begin{aligned}\{\text{Current} &= [0.2272, 0.2408], \text{Active Power}\\ &= [38.5260, 39.9682], \text{Reactive Power}\\ &= [42.1306, 44.8896]\} \Rightarrow\\ \{\text{Voltage} &= [142.3809, 144.0914)\}\end{aligned}$$

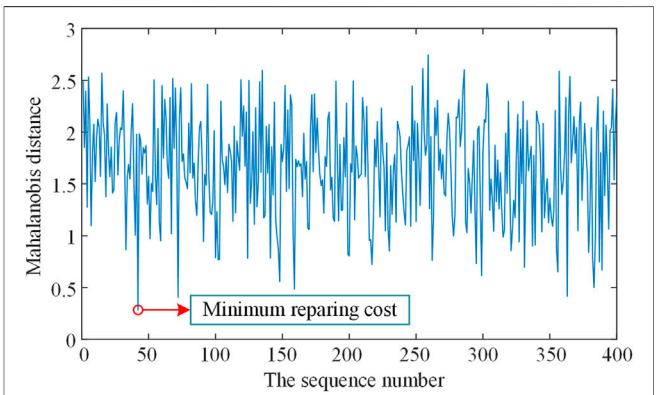

FIGURE 8 | The repairing cost of one outlier.

TABLE 1 | Confusion matrix of outliers detection.

Actual label	Detection results	
	Outlier/1	Normal/0
Outlier/1	TP	FN
Normal/0	FP	TN

According to this rule, the correct value of voltage magnitude should be in the range of [142.3809,144.0914). It means that after the outliers are repaired, the point is still in the original sub-RBS. This is because data repairing can be translated into searching for normal data with the highest similarity to it within the RBS.

The Mahalanobis distance is used as a metric to account for the distributional differences between attributes. The repair results within this RBS are not unique, and each result has a corresponding repair cost. As shown in **Figure 8**, in outlier repairing, the objective is to minimize the repairing cost function as **Eq. 8**.

$$\boldsymbol{cost}\left(x, x'\right) = \boldsymbol{M}_d\left(x, x'\right) \tag{8}$$

where x' is the repairing result of the corresponding x.

However, in some cases, the minimum value of the cost function is more than one. In these cases, the abnormal observation will be replaced by O_t', which calculated by the following **Eq. 9**.

$$\boldsymbol{O}_t' = \boldsymbol{mean}\left(\boldsymbol{D}_{C_{\min}}\right) \tag{9}$$

where $D_{C_{\min}} \in \boldsymbol{D}$ that corresponds to C_{min}, and C_{min} is the set with the minimize repairing cost.

In addition, we consider the outliers repairing with the data feedback. When an outlier is cleaned to a normal value after outliers repairing, the observation would be updated in the RBS for a new outlier that needs repair. The accuracy of outliers repairing could be improved with data feedback.

EXPERIMENT AND ANALYSIS

The Metrics Used for Evaluating

Outliers detection of measurement data is an unbalanced binary classification problem. Data are classified as normal or abnormal. In this way, accuracy is not an appropriate metric for evaluating the performance of a method. The detection results could be classified into four types according to the label between actual and prediction values: true positive (TP), true negative (TN), false positive (FP) and false negative (FN). The confusion matrix of outliers detection is shown in **Table 1**. In general, the *Precision*, *Recall* and *F1-Score* are used as the metrics for evaluating of classification problem. Among them, *the Precision* is a metrics reflects the reliability of the detection results, while *the Recall* is a metrics reflects how many truly detection results are returned. And the *F1-Score* is the harmonic mean of precision and recall.

According to the confusion matrix of outliers detection, the *Precision*, *Recall* and *F1-Score* could be calculated by **Eqs 10–12**.

$$\boldsymbol{Precision} = \frac{\boldsymbol{TP}}{\boldsymbol{TP} + \boldsymbol{FP}} \tag{10}$$

$$\boldsymbol{Recall} = \frac{\boldsymbol{TP}}{\boldsymbol{TP} + \boldsymbol{FN}} \tag{11}$$

$$\boldsymbol{F1 - Score} = \frac{\boldsymbol{2 \times Precision \times Recall}}{\boldsymbol{Precision + Recall}} \tag{12}$$

where TP is the count of outlier detected as an outlier, FP is the count of normal data detected as an outlier, FN is the count of outlier detected as normal data.

In addition, two metrics used for outliers repairing: the mean absolute error (MAE) and the root mean square error (RSME). They are defined as follow **Eqs 13**, **14** .

$$\boldsymbol{MAE} = \frac{\boldsymbol{1}}{\boldsymbol{N}} \sum_{i=1}^{N} \left| \hat{x}_i - x_j' \right| \tag{13}$$

$$\boldsymbol{RMSE} = \frac{\boldsymbol{1}}{\boldsymbol{N}} \sqrt{\sum_{i=1}^{N} \left(\hat{x}_i - x_i' \right)^2} \tag{14}$$

where N is the size of data, $\hat{x}_i$ is the ith actual value (without contaminated) of outliers, x_i' is the estimation of the ith outlier.

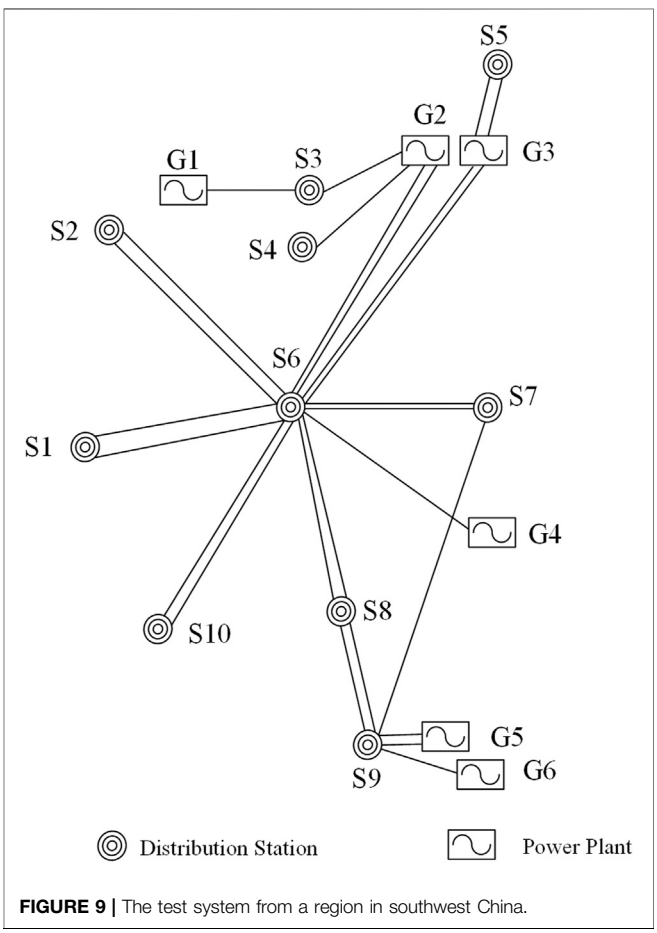

FIGURE 9 | The test system from a region in southwest China.

TABLE 2 | The noises injection of simulated dataset.

Anomaly rate	The outliers calculation in each feature	Anomaly proportion
Noise 5%	1.p.u *105% + G(x)	569/4000
Noise 10%	1.p.u *105% + G(x)	1091/4000
Noise 15%	1.p.u *105% + G(x)	1529/4000

The Simulated Dataset

Unfortunately, the measurement datasets from real-world distribution networks are unlabeled. It means that it is not appropriate to use as a dataset for evaluating the proposed methods. Hence, we used the simulated datasets with artificial error. To verify the correctness and effectiveness of the proposed method, a test system (from a region in southwest China) is modeled in PSCAD/EMTDC to collect simulation data, as shown in **Figure 9**. The operational datasets contain 4000 samples (with a sampling rate of 40 frames per second) and four features (voltage magnitude V, current magnitude I, active power P, reactive power Q) for the distribution network. There are no outliers in these datasets. We added some synthetic errors to the simulated measurement data, in which outliers are generated and injected into the datasets using a normal-distributed random function as $z = G(x)$. **Table 2** shows that each bus data has 5–15% noise injected into it. As an example, if a data point has a voltage magnitude feature of 110*k*V, the noise is calculated as 110*105% + G(x). For each sample, three-fourths of the data is taken as input, and the trained algorithm predicts the rest of the real-time value.

Outliers Detection

The discretization of the pre-processed datasets is performed using k-means clustering. The numerical attributes voltage magnitude, current magnitude, active power and reactive power are clustered into 8, 5, 6, and 5 categories, respectively. The clusters are selected based on the quality metric that is finally estimated. After data discretization, the Apriori algorithm generates the association rules in the test datasets with confidence greater than 60%. For each posterior feature, the association rules are generated separately. Then the rules are generated for prediction individuals. Some of these rules are shown in **Table 3**. If the observation is not within the interval determined by the rules, it will be marked as an outlier.

For evaluation, the proposed method is compared with other methods such as decision tree, k-neighbors, and SVM; all methods are using simulated datasets. The results are shown in **Table 4** and **Figure 10**, and the comparison is based on *Precision*, *Recall* and *F1-Score*, respectively. For the *Precision*, considering the dataset with 5–15% noise, the above method have similar results, which means that the change of anomaly rate has little effect on the Precision. For the *Recall*, with the anomaly rate increases, the above method have worse results, which means that the change of anomaly rate mainly affects the *Recall*, resulting in the change of *F1-score*.

According to **Table 4** and **Figure 10**, the result of decision tree is not satisfactory, which may mistakenly treats the normal data as an outlier. And for the datasets with highly contaminated (more than 15%), the SVM is leaving much to be desired, the *Recall* of it even less than 90%. For SVM, the reason may be that the high anomaly rate makes the training data extremely unbalanced. Therefore, in the training stage, the type of data may not meet the requirements of SVM, which limits the application of SVM in outliers detection. Under the same conditions, the *Precision* and *Recall* of k-neighbors is small than our proposed method. From the **Figure 10**, the proposed method play a good performance, whose *F1-Score* remains more than 96% for the datasets with different anomaly rate. The comparative case studies show that our proposed method outperformed the other three methods.

TABLE 3 | The part of association rules in different anomaly rate (voltage as posterior).

Anomaly rate (%)	Prior	Posterior	*Con*
5	{Current = [0.2408, 0.2521), Active Power = [38.5260, 39.9682), Reactive Power = [45.6303, 46,6837)}	{Voltage = [142.3809, 144.0914)}	0.7692
	{Current = [0.2272, 0.2408], Active Power = [38.5260, 39.9682), Reactive Power = [41.1306, 43.8896]}	{Voltage = [134.5451, 137.8775]}	0.7222
	{Current = [0.2272, 0.2408], Active Power = [38.0519, 38.5260), Reactive Power = [45.6303, 46,6837)}	{Voltage = [144.0914, 149.0609)}	0.6818
	{Current = [0.2272, 0.2408], Active Power = [38.0519, 38.5260), Reactive Power = [41.1306, 43.8896]}	{Voltage = [137.8775, 140.3127)}	0.6800
	{Current = [0.2408, 0.2521), Active Power = [38.5260, 39.9682), Reactive Power = [43.8896, 45.6303)}	{Voltage = [140.3127, 142.3809)}	0.6800
	...	...	...
10	{Current = [0.2272, 0.2408], Active Power = [36.9662, 38.0519], Reactive Power = [46.6837, 48.0561)}	{Voltage = [144.0914, 149.0609)}	0.7826
	{Current = [0.2408, 0.2521), Active Power = [38.0519, 38.5260), Reactive Power = [45.6303, 46,6837)}	{Voltage = [142.3809, 144.0914)}	0.7407
	{Current = [0.2408, 0.2521), Active Power = [38.0519, 38.5260), Reactive Power = [45.6303, 46,6837)}	{Voltage = [137.8775, 140.3127)}	0.7333
	{Current = [0.2272, 0.2408], Active Power = [38.5260, 39.9682), Reactive Power = [45.6303, 46,6837)}	{Voltage = [134.5451, 137.8775]}	0.7142
	{Current = [0.2272, 0.2408], Active Power = [39.9682, 42.3351), Reactive Power = [41.1306, 43.8896]}	{Voltage = [140.3127, 142.3809)}	0.6785
	...	...	...
15	{Current = [0.2272, 0.2408], Active Power = [38.5260, 39.9682), Reactive Power = [43.8896, 45.6303)}	{Voltage = [140.3127, 142.3809)}	0.8181
	{Current = [0.2272, 0.2408], Active Power = [39.9682, 42.3351), Reactive Power = [46.6837, 48.0561)}	{Voltage = [137.8775, 140.3127)}	0.7368
	{Current = [0.2272, 0.2408], Active Power = [36.9662, 38.0519], Reactive Power = [45.6303, 46,6837)}	{Voltage = [134.5451, 137.8775]}	0.7333
	{Current = [0.2408, 0.2521), Active Power = [38.5260, 39.9682), Reactive Power = [45.6303, 46,6837)}	{Voltage = [142.3809, 144.0914)}	0.7000
	{Current = [0.2408, 0.2521), Active Power = [38.5260, 39.9682), Reactive Power = [46.6837, 48.0561)}	{Voltage = [144.0914, 149.0609)}	0.6957
	...	...	...

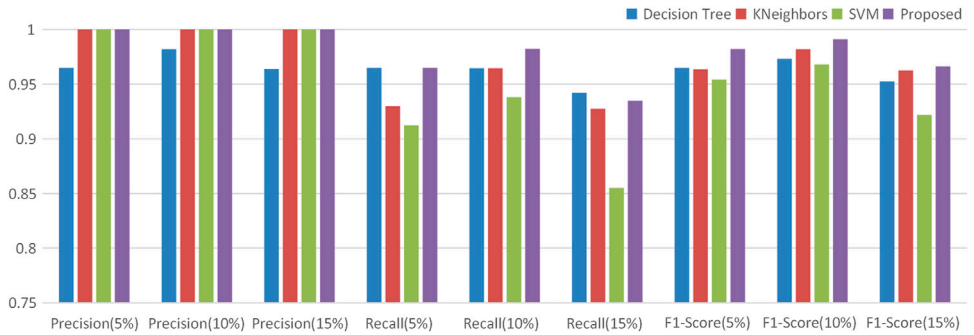

FIGURE 10 | The histogram of the comparative analysis in different outliers detection methods.

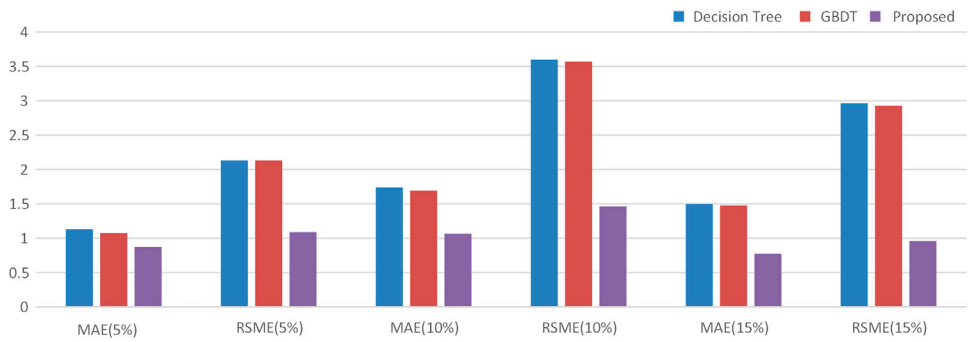

FIGURE 11 | The histogram of the comparative analysis in different outliers repairing methods.

TABLE 4 | The results of the comparative analysis in different outliers detection method.

Anomaly rate (%)	Method	Precision	Recall	F1-score
5	Decision Tree	0.9649	0.9649	0.9649
	K-Neighbors	1.00	0.9298	0.9636
	SVM	1.00	0.9123	0.9541
	Proposed	1.00	0.9649	0.9821
10	Decision Tree	0.9820	0.9646	0.9732
	K-Neighbors	1.00	0.9646	0.9820
	SVM	1.00	0.9381	0.9680
	Proposed	1.00	0.9823	0.9911
15	Decision Tree	0.9630	0.9420	0.9524
	K-Neighbors	1.00	0.9275	0.9624
	SVM	1.00	0.8551	0.9219
	Proposed	1.00	0.9348	0.9663

TABLE 5 | The results of the comparative analysis in different outliers repairing methods.

Anomaly rate (%)	Method	MAE	RSME
5	Decision Tree	1.1307	2.1325
	GBDT	1.0744	2.1315
	Proposed	0.8720	1.0871
10	Decision Tree	1.7386	3.5983
	GBDT	1.6941	3.5701
	Proposed	1.0647	1.4617
15	Decision Tree	1.5002	2.9637
	GBDT	1.4785	2.9260
	Proposed	0.7751	0.9595

Outliers Repairing

We employ two other widely-used data repairing methods to make a comparison, which includes decision tree and gradient boosting decision tree (GBDT). **Table 5** and **Figure 11** show the comparison result of the MAE and RSME at different anomaly rates. The accuracy indices of the decision tree and GBDT is close under different anomaly rates. And the GBDT algorithm is in a better position than the decision tree algorithm for each evaluation indices. The proposed method has been superior in all the accuracy indices in terms of aggregate, indicating model robustness.

In these cases, we only showed the result for detection and repairing the voltage, but the proposed method can also work for other features. In general, the proposed method outperforms the other methods in most cases. However, our approach's performance may not be good in some situations, especially when used with little historical data. The reason for the problem is the insufficiency of association rules. The

association rules can not include all the conditions, which cause by the value of confidence (more than 60%). One way to improve the accuracy of the proposed method is to increase the amount of historical data. So that more association rules can be generated to mine the correlation between features from the data.

CONCLUSION

In this paper, we developed a association rules-based method for outliers cleaning. To detect outliers, the association rules are generated from historical data in conjunction with DBSCAN, k-means and Apriori technique. For the outliers repairing, we took into account the repairing cost by a distance-based model. The Mahalanobis distance was used for constructing a data repairing cost function to reduce the errors. The proposed method achieves accurate detection as compared to decision tree, k-neighbors, and SVM algorithms. When outliers is taken into account, our model produces a smaller MAE and RSME, which has a better result than decision tree and GBDT. The results show that the work has a positive effect on improving data quality, which means our works could provide a reliable data base for distribution network planning and operation. Future work will focus on combining this approach with Spark parallel computing technology to improve the efficiency of the algorithm to satisfy the practical application needs of distribution network measurement outliers cleaning.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusion of this article will be made available by the authors, without undue reservation.

AUTHOR CONTRIBUTIONS

Conception and design of study: HK; Acquisition of data: MH, XH; Drafting the article: RQ, XH; Analysis and interpretation of data: HK, RQ, CG, and XM; Revising the article critically for important intellectual content: RQ, XM, and RD.

FUNDING

This work was supported by the Science and Technology Foundation of China Southern Power Grid (YNKJXM20191369).

REFERENCES

Alimardani, A., Therrien, F., Atanackovic, D., Jatskevich, J., and Vaahedi, E. (2015). Distribution System State Estimation Based on Nonsynchronized Smart Meters. *IEEE Trans. Smart Grid* 6 (6), 2919–2928. doi:10.1109/TSG.2015.2429640

Cai, Y., Xiao, X., Tian, H., Fu, Y., Wu, P., and He, H. (2021). A Multi-Source Data Collection and Information Fusion Method for Distribution Network Based on Iot Protocol. *IOP Conf. Ser. Earth Environ. Sci.* 651 (2), 022076. doi:10.1088/1755-1315/651/2/022076

Chen, J., Li, W., Lau, A., Cao, J., and Wang, K. (2010). Automated Load Curve Data Cleansing in Power Systems. *IEEE Trans. Smart Grid* 1 (2), 213–221. doi:10.1109/TSG.2010.2053052

Chen, K., Jiang, Y., Wu, Z., Zheng, N., Wang, H., and Hong, H. (2021). HTsort: Enabling Fast and Accurate Spike Sorting on Multi-Electrode Arrays. *Front. Comput. Neurosci.* 15, 657151. doi:10.3389/fncom.2021.657151

Chengyu, C., and Ying, X. (2016). Research and Improvement of Apriori Algorithm for Association Rules. *Phys. Rev. A*, 1–4. doi:10.1103/PhysRevA.94.042311

Chipade, V. S., Marella, V. S. A., and Panagou, D. (2021). Aerial Swarm Defense by StringNet Herding: Theory and Experiments. *Front. Robot. AI* 8, 640446. doi:10.3389/frobt.2021.640446

Esmalifalak, M., Liu, L., Nguyen, N., Zheng, R., and Han, Z. (2014). Detecting Stealthy False Data Injection Using Machine Learning in Smart Grid. *IEEE Syst. J.*, 11, 1–9. doi:10.1109/JSYST.2014.2341597

Hayes, B. P., Gruber, J. K., and Prodanovic, M. (2018). Multi-nodal Short-term Energy Forecasting Using Smart Meter Data. *IET Generation, Transm. Distribution* 12 (12), 2988–2994. doi:10.1049/iet-gtd.2017.1599

Hu, X., Zhang, H., Ma, D., and Wang, R. (2021). Hierarchical Pressure Data Recovery for Pipeline Network via Generative Adversarial Networks. *IEEE Trans. Automat. Sci. Eng.* (99), 1–11. doi:10.1109/TASE.2021.3069003

Li, X., Cai, Y., and Zhu, W. "Power Data Cleaning Method Based on Isolation Forest and LSTM Neural Network," in International Conference on Cloud Computing and Security, Haikou, China, 26 September 2018, 539–550. doi:10.1007/978-3-030-00018-9_47

Liu, J., Cao, Y., Li, Y., Guo, Y., and Deng, W. (2020). A Big Data Cleaning Method Based on Improved CLOF and Random Forest for Distribution Network. *CSEE J. Power Energy Syst.* (Early Access). doi:10.17775/CSEEJPES.2020.04080

Liu, S., Zhao, Y., Lin, Z., Ding, Y., Yan, Y., Yang, L., et al. (2019). Data-Driven Condition Monitoring of Data Acquisition for Consumers' Transformers in Actual Distribution Systems Using T-Statistics. *IEEE Trans. Power Deliv.* 34 (4), 1578–1587. doi:10.1109/TPWRD.2019.2912267

Liu, X., Zhang, X., Chen, L., Xu, F., and Feng, C. (2020). Data-driven Transient Stability Assessment Model Considering Network Topology Changes via Mahalanobis Kernel Regression and Ensemble Learning. *J. Mod. Power Syst. Clean Energ.* 8 (6), 1080–1091. doi:10.35833/MPCE.2020.000341

Maesschalck, R. D., Jouan-Rimbaud, D., and Massart, D. L. (2000). The Mahalanobis Distance. *Chemometrics Intell. Lab. Syst.* 50 (1), 1–18. doi:10.1016/S0169-7439(99)00047-7

Mccamish, B., Meier, R., Landford, J., Bass, R. B., Chiu, D., and Cotilla-Sanchez, E. (2016). A Backend Framework for the Efficient Management of Power System Measurements. *Electric Power Syst. Res.* 140 (nov), 797–805. doi:10.1016/j.epsr.2016.05.003

Nascimento, R. M. D., Oening, A. P., Marcilio, D. C., Alexandre, R. A., Júnior, E. P. R., and Schiochet, J. M. ""Outliers' Detection and Filling Algorithms for Smart Metering Centers"," in Proceedings of the 2012 IEEE PES Transmission and Distribution Conference and Exposition, Orlando, Florida, USA, May 2012, 7–10. doi:10.1109/tdc.2012.6281659

Nemati, H., Laso, A., Manana, M., Sant'Anna, A., and Nowaczyk, S. (2018). Stream Data Cleaning for Dynamic Line Rating Application. *Energies* 11 (8). doi:10.3390/en1101200710.3390/en11082007

Pei, Z., Li, F., and Bhatt, N. (2010). Next-generation Monitoring, Analysis, and Control for the Future Smart Control center. *IEEE Trans. Smart Grid* 1 (2), 186–192. doi:10.1109/TSG.2010.2053855

Qu, Z. Y., Wang, Y. W., Wang, C., Qu, N., and Yan, J. (2016). A Data Cleaning Model for Electric Power Big Data Based on Spark Framework. *Adv. Sci. Technology* 9, 137–150. doi:10.14257/astl.2016.121.74

Rauch, J. (2005). Logic of Association Rules. *Appl. Intelligence* 22 (1), 9–28. doi:10.1023/B:APIN.0000047380.15356.7a

Shi, X., Qiu, R., Ling, Z., Yang, F., Yang, H., and He, X. (2020). Spatio-Temporal Correlation Analysis of Online Monitoring Data for Anomaly Detection and

Location in Distribution Networks. *IEEE Trans. Smart Grid* 11 (2), 995–1006. doi:10.1109/TSG.2019.2929219

Song, Y., Zhou, G., and Zhu, Y. (2013). Present Status and Challenges of Big Data Processing in Smart Grid. *Power Syst. Technology* 37 (4), 927–938. doi:10.3969/j.issn.1006-9402.2014.05.038

Thams, F., Venzke, A., Eriksson, R., and Chatzivasileiadis, S. (2020). Efficient Database Generation for Data-Driven Security Assessment of Power Systems. *IEEE Trans. Power Syst.* 35 (1), 30–41. doi:10.1109/TPWRS.2018.2890769

Thang, T. M., and Kim, J. The Anomaly Detection by Using DBSCAN Clustering with Multiple Parameters." in Proceedings of the 2011 International Conference on Information Science and Applications, Jeju, Korea (South), April 2011, IEEE, 1–5. doi:10.1109/ICISA.2011.5772437

Waal, T. D., Pannekoek, J., and Scholtus, S. (2011). *Handbook of Statistical Data Editing and Imputation*. Hoboken, New Jersey, USA: John Wiley & Sons. ISBN: 978-0-470-54280-4.

Wang, Q., Li, F., Tang, Y., and Xu, Y. (2019). Integrating Model-Driven and Data-Driven Methods for Power System Frequency Stability Assessment and Control. *IEEE Trans. Power Syst.* 34 (6), 4557–4568. doi:10.1109/TPWRS.2019.2919522

Wang, R., Sun, Q., Tu, P., Xiao, J., Gui, Y., and Wang, P. (2021). Reduced-order Aggregate Model for Large-Scale Converters with Inhomogeneous Initial Conditions in Dc Microgrids. *IEEE Trans. Energ. Convers.* 36 (99), 2473–2484. doi:10.1109/TEC.2021.3050434

Wang, Y., Chen, Q., Hong, T., and Kang, C. (2018). Review of Smart Meter Data Analytics: Applications, Methodologies, and Challenges. *IEEE Trans. Smart Grid*, 10, 1. doi:10.1109/TSG.2018.2818167

Yan, J. Z., Gao, Y., and Yu, Y. C. "Water Quality Data Outlier Detection Method Based on Spatial Series Features," in Proceedings of the The 6th International Conference on Fuzzy Systems and Data Mining (FSDM), Xiamen, China, November 2020. doi:10.3233/FAIA200715

Yan, Y., Sheng, G., Chen, Y., Jiang, X., and Du, X. (2015). An Method for Anomaly Detection of State Information of Power Equipment Based on Big Data Analysis. *Proc. Csee* 35 (1), 52–59. doi:10.13334/j.0258-8013.pcsee.2015.01.007

Ye, Y., Wang, Z. D., Zhang, Z. Y., Zhao, J. G., and Zhai, L. (2010). An Estimation Method of Energy Loss for Distribution Network Planning. *Power Syst. Prot. Control.* 17, 82–86. doi:10.3969/j.issn.1674-3415.2010.17.016

Conflict of Interest: Author HK is employed by Yunnan Power Grid Co., Ltd., China.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Energy Research

ORIGINAL RESEARCH
published: 20 October 2021
doi: 10.3389/fenrg.2021.750983

Wind Turbine Pitch System Fault Detection Using ssODM-DSTA

Mingzhu Tang[1†], Jiahao Hu[1], Huawei Wu[2] and Zimin Wang[3†]*

[1]School of Energy and Power Engineering, Changsha University of Science and Technology, Changsha, China, [2]Hubei Key Laboratory of Power System Design and Test for Electrical Vehicle, Hubei University of Arts and Science, Xiangyang, China, [3]School of Computer Science and Information Security, Guilin University of Electronic Technology, Guilin, China

A fault detection method of wind turbine pitch system using semi-supervised optimal margin distribution machine (ssODM) optimized by dynamic state transition algorithm (DSTA) [ssODM-DSTA] was proposed to solve the problem of obtaining the optimal hyperparameters of the fault detection model for the pitch system. This method was adopted to input the three hyperparameters of the ssODM into the dynamic state transition algorithm in the form of a three-dimensional vector to obtain the global optimal hyperparameters of the model, thus improving the performance of the fault detection model. Using a random forest to rank the priority of features of the pitch system fault data, the features with large weight proportions were retained. Then, the Pearson correlation method is used to analyze the degree of correlation among features, filter redundant features, and reduce the scale of features. The dataset was divided into a training dataset and a test dataset to train and test the proposed fault detection model, respectively. The real-time wind turbine pitch system fault data were collected from domestic wind farms to carry out fault detection experiments. The results have shown that the proposed method had a positive fault rate (FPR) and fault negative rate (FNR), compared with other optimization algorithms.

Keywords: fault detection, wind turbine, pitch system, dynamic state transition algorithm, semi-supervised optimal margin distribution machine, random forest

OPEN ACCESS

Edited by:
Qiuye Sun,
Northeastern University, China

Reviewed by:
Xuguang Hu,
Northeastern University, China
Franklin Chang,
Nanyang Technological University, Singapore

****Correspondence:***
Huawei Wu
whw_xy@hbuas.edu.cn

†These authors have contributed equally to this work

Specialty section:
This article was submitted to Smart Grids, a section of the journal Frontiers in Energy Research

Received: *31 July 2021*
Accepted: *17 August 2021*
Published: *20 October 2021*

Citation:
Tang M, Hu J, Wu H and Wang Z (2021) Wind Turbine Pitch System Fault Detection Using ssODM-DSTA. Front. Energy Res. 9:750983. doi: 10.3389/fenrg.2021.750983

INTRODUCTION

China's energy structure is unceasingly transforming towards low carbon and environmental protection by striving to build a new energy system and vigorously advocating the development of renewable energy industry to achieve the overall objective of carbon emission peak by 2030 and carbon neutrality by 2060 (Li and Bo, 2020; Qian and Wang, 2020; Sun et al., 2015). Renewable energy is a kind of clean and green energy that can replace traditional energy. The emergence of this energy has made a great contribution to reducing pollution and emission in the world. Renewable energy power systems (Zhang et al., 2021; Zhang and Ruan, 2019) mainly include wind power generation, solar power generation, and hydropower generation. Wind power, as an indispensable part of renewable energy, continues to expand in scale. By the end of 2020, China's installed capacity of grid-connected wind power has reached 281.53 million kW, increasing by 34.6% year over year, accounting for 12.79% of the total installed capacity (Blaabjerg et al., 2012; Blaabjerg and Ke Ma, 2013). However, as the wind power industry is rapidly developing, the maintenance and repair pressure of wind turbines is also increasing. The operating environment of wind turbines is relatively bad. Abnormal climate, unstable wind speed, and other factors often lead to the faults of wind turbines and the shutdown (Song et al., 2021; Yang et al., 2021). The fault rate and complexity of the

pitch system, an important part of the wind turbine, are higher than those of the main shaft, gearbox, and generator (Wang et al., 2021). In case of any fault, the power generation rate of the wind turbine will be directly affected, leading to damage to the wind turbine and huge economic losses. For this reason, effective fault detection is of great significance for wind turbine pitch systems.

In the wake of the era of big data and emerging machine learning, fault detection and fault diagnosis methods for wind turbines based on machine learning algorithms have gradually matured in recent years (Hu et al., 2021; Tang et al., 2021; Long et al., 2020). With the wind turbine fault data in the SCADA system, which can collect, monitor, and control the operation data of wind turbines in real-time, it is a common and reliable method to choose an appropriate machine learning algorithm for the fault detection of wind turbines. Machine learning algorithms mainly include supervised learning, unsupervised learning, semi-supervised learning, and reinforcement learning. Generally, classification methods for wind turbine fault detection include support vector machine (SVM), artificial neural networks (ANNs) (Xi et al., 2021), and Large margin distribution machine (LDM) (Zhang and Zhou, 2014). In the literature (Tuerxun et al., 2021), a SVM based on a sparrow search algorithm (SSA) was used for wind turbine fault diagnosis and had achieved excellent results. Moreover, a fault diagnosis method based on stochastic subspace identification and multinuclear SVM was proposed to identify bearing faults of wind turbines (Zhao et al., 2019). The ANNs and empirical mode decomposition (EMD) were used to effectively identify different turbine imbalance faults in (Malik and Mishra, 2017). In (Zhang and Wang, 2014), the artificial neural networks had a great diagnosis effect on the main bearing of the wind turbine during early fault prediction. The cost-sensitive large margin distribution machine (CLDM) proposed in (Tang et al., 2019) can better deal with the classification imbalance data and misclassification cost inequality of wind turbine datasets.

The above methods are only applicable to the input data with characteristic values and tags. However, tags are usually scarce and expensive in the actual wind turbine data, which can be effectively dealt with using the semi-supervised learning method. Commonly used semi-supervised learning methods include transductive support vector machine (TSVM), safe semi-supervised support vector machine (S4VM), and Laplacian support vector machine (LaSVM) (Chong et al., 2020). In (Shen et al., 2012), a gear reducer fault diagnosis model based on EMD and TSVM was proposed to solve the problem of insufficient tags of gear reducer data samples, and results have shown a high fault diagnosis accuracy. In (Li, 2010), graph theory and transductive support vector machine (GTSVM) was used to solve the problem of insufficient fault samples for training in mechanical fault diagnosis, and results have shown that this method improved the accuracy of fault diagnosis. A new fault alarm rule based on the upper bound of S4VM generalized error proposed in (Mao et al., 2020) can self-adaptively identify the occurrence of early bearing faults. In (Dai et al., 2017), a rolling bearing fault diagnosis method based on composite multi-scale entropy (CMSE), sequential forward modeling selection, and LaSVM was proposed to solve the problem of the large sample size of and tagging difficulty in rolling bearing fault diagnosis, and results have shown that the effect of fault diagnosis was improved.

The semi-supervised optimal margin distribution machine (Zhang and Zhou, 2018) was a classification algorithm with high generalization ability, proposed for generalization ability based on optimal margin distribution machine (ODM) (Tan et al., 2020). "Lables" were given to the samples without lables and the semi-supervised learning was transformed into "supervised learning" *via* this algorithm. On this basis, on the premise of optimizing the minimum margin and maximizing the hyperplane, the distribution of sample margin was fully considered and the mean value and variance between samples were introduced to improve the classification ability of the algorithm.

The reasonable selection of hyperparameters can significantly affect the fault detection performance during fault detection for wind turbine pitch systems based on a machine learning algorithm. For this reason, the optimal hyperparameters of the fault detection model should be obtained by optimizing the parameter optimization algorithm (Long et al., 2021a; 2021b). In (Zhang et al., 2020), a particle swarm optimization algorithm (PSO) was used to optimize SVM for fault diagnosis of wind turbine gearbox bearings, and results have shown that the precision and accuracy of diagnosis were improved. In (Chen, 2020), backpropagation neural network (BPNN) and long short-term memory network (LSTMN) were combined with PSO and great fault diagnosis results were obtained in wind turbine rolling bearing fault diagnosis. In (Odofin et al., 2018), a genetic algorithm (GA) was adopted to optimize the machine learning algorithm to improve the reliability of the wind turbine energy system. In (Zhang et al., 2018), GA was introduced into anomaly identification of wind turbine state parameters to successfully optimize the anomaly identification results. In (Yao et al., 2021), grid search (GS) was used to optimize the fault classification algorithm during battery fault diagnosis and the fault diagnosis accuracy was improved. In (Zhang and Sheng, 2021), GS was used to optimize the hyperparameters and kernel functions of support vector machines to improve the accuracy of the motor fault diagnosis.

With a design based on a state transition algorithm (STA) (Zhou et al., 2012), the DSTA (Zhou et al., 2018) is a dynamic stochastic intelligent global optimization method with its own risk p_{risk} and restoration in probability p_{rest} adjustment strategy. For the fault detection model of wind turbine pitch system with high complexity, using some of the above-mentioned common optimization algorithms to optimize its hyperparameters is often easy to fall into local optimization because the scale of optimization object is too large and the optimization problem is too complex. Facing these problems, DSTA can use a dynamic adjustment strategy to surpass local optimization, and the optimization algorithm provides four search operators and novel update and selection methods to support its excellent searchability according to the different needs of optimization objectives. Using DSTA to optimize the fault detection model of the wind turbine pitch system can carry out global search and converge quickly. It is a novel optimal combination of fault detection models.

For complex and variable pitch system faults, it is often difficult to select the optimal parameters for the fault detection model of the wind turbine pitch system. Meanwhile, variable pitch fault data without tags will lead to unsatisfactory fault detection results. For this problem, a method of optimizing a ssODM based on a DSTA was proposed.

SEMI-SUPERVISED OPTIMAL MARGIN DISTRIBUTION MACHINE

Suppose the mean margin value of training dataset samples after normalization is $\widehat{r}_m$. The difference between the margin of the sample (x_i, y_i) after normalization and the mean margin value was $\left|\widehat{\gamma}(x_i, y_i) - \widehat{\gamma}_m\right|$. So the variance between the maximum mean margin value and the minimum mean margin value could be represented in the following form:

$$\begin{aligned} &\max_{\omega,\widehat{\xi}_i,\widehat{\epsilon}_i} \rho\widehat{\gamma}_m^2 - \frac{1}{m}\sum_{i=1}^{m}\left(\widehat{\xi}_i^2 + \widehat{\epsilon}_i^2\right) \\ &s.t.\widehat{\gamma}\left(x_i, y_i\right) \geq \widehat{\gamma}_m - \widehat{\xi}_i \\ &\widehat{\gamma}\left(x_i, y_i\right) \leq \widehat{\gamma}_m + \widehat{\epsilon}_i, \forall i \in [m], \end{aligned} \tag{1}$$

where the parameter ρ was used to weigh two priorities. As $\widehat{\gamma}(x_i, y_i)$ could not be more than and less than the mean value at the same time, there were a nonnegative value and a non-zero value in $\widehat{\xi}_i, \widehat{\epsilon}_i$. The second item of the objective function was the margin variance. $\xi_i = \omega\widehat{\xi}_i$ and $\epsilon_i = \omega\widehat{\epsilon}_i$. The above equation could be rewritten as follows:

$$\begin{aligned} &\max_{\omega,\xi_i,\epsilon_i} \rho\frac{\widehat{\gamma}_m^2}{\omega^2} - \frac{1}{m}\sum_{i=1}^{m}\frac{\xi_i^2 + \epsilon_i^2}{\omega^2} \\ &s.t.y_i\omega^T\phi\left(x_i\right) \geq \gamma_m - \xi_i \\ &y_i\omega^T\phi\left(x_i\right) \leq \gamma_m + \epsilon_i, \forall_i \in [m], \end{aligned} \tag{2}$$

where γ_m did not affect the optimization. When γ_m was scaled, ω, ξ_i *and* ϵ_i were scaled on the same scale. In this case, the constraint was still satisfied, and the objective function value remained unchanged. Set $\gamma_m = 1$; the equation could be further rewritten as follows:

$$\max_{\omega,\xi_i,\epsilon_i} \frac{1}{\omega^2}\left(1 - \frac{1}{m}\sum_{i=1}^{m}\xi_i^2 + \epsilon_i^2\right) s.t.y_i\omega^T\phi\left(x_i\right) \geq 1 - \xi_i y_i\omega^T\phi\left(x_i\right) \leq 1 + \epsilon_i, \forall_i \in [m]. \tag{3}$$

As the maximum objective function was equal to the minimum ω^2 and $\sum_{i=1}^{m}\xi_i^2 + \epsilon_i^2$, there was a constant λ to make the above optimization equation have the same solution as the following equation:

$$\begin{aligned} &\max_{\omega,\xi_i,\epsilon_i} \frac{1}{2}\omega^2 + \frac{\lambda}{m}\sum_{i=1}^{m}\left(\xi_i^2 + \epsilon_i^2\right) \\ &s.t.y_i\omega^T\phi\left(x_i\right) \geq 1 - \xi_i \\ &y_i\omega^T\phi\left(x_i\right) \leq 1 + \epsilon_i, \forall_i \in [m]. \end{aligned} \tag{4}$$

A parameter $\mu \in (0, 1)$ was introduced to weigh the deviation loss in two different directions between the sample margin and the mean margin value 1. A parameter θ-insensitive loss was introduced to control model sparsity. Thus, the final equation form was as follows:

$$\begin{aligned} &\max_{\omega,\xi_i,\epsilon_i} \frac{1}{2}\omega^2 + \frac{\lambda}{m}\sum_{i=1}^{m}\frac{\left(\xi_i^2 + \mu\epsilon_i^2\right)}{(1-\theta)^2} \\ &s.t.y_i\omega^T\phi\left(x_i\right) \geq 1 - \theta - \xi_i \\ &y_i\omega^T\phi\left(x_i\right) \leq 1 + \theta + \epsilon_i, \forall_i \in [m]. \end{aligned} \tag{5}$$

$\widehat{\boldsymbol{y}} = [\widehat{y}_i, \ldots, \widehat{y}_m \in \{\pm 1\}^m]$ was unlabeled. The ssODM could be converted into the following form:

$$\begin{aligned} &\min_{\widehat{y}\in\mathcal{B}} \min_{\omega,\xi_i,\epsilon_i} \frac{1}{2}\omega^2 + \frac{\lambda_1}{l}\sum_{i=1}^{l}\frac{\xi_i^2 + \mu\epsilon_i^2}{(1-\theta)^2} + \frac{\lambda_2}{u}\sum_{i=l+1}^{l+u}\frac{\xi_i^2 + \mu\epsilon_i^2}{(1-\theta)^2} \\ &s.t.y_i\omega^T\phi\left(x_i\right) \geq 1 - \theta - \xi_i \\ &y_i\omega^T\phi\left(x_i\right) \leq 1 + \theta + \epsilon_i, \forall_i \in [m]. \end{aligned} \tag{6}$$

where $\mathcal{B} = \left\{\widehat{\boldsymbol{y}}\middle|\frac{e^T\widehat{y}_U}{m-l} = \frac{e^T y_L}{l}\right\}$ was equilibrium constraint to prevent the occurrence of trivial solutions. $\lambda_i = \frac{\lambda_1(m-l)-\lambda_2 l}{l(m=l)}1_{i\in L} + \frac{\lambda_2}{m-l}$. λ_1 a*nd* λ_2 were loss parameters to weigh the tagged and untagged data. The above questions could be further written as follows:

$$\begin{aligned} &\min_{\widehat{y}\in\mathcal{B}} \min_{\omega,\xi_i,\epsilon_i} \frac{1}{2}\omega^2 + \sum_{i=1}^{m}\lambda_i\frac{\xi_i^2 + \mu\epsilon_i^2}{(1-\theta)^2} \\ &s.t.\widehat{y}_i\omega^T\phi\left(x_i\right) \geq 1 - \theta - \xi_i \\ &\widehat{y}_i\omega^T\phi\left(x_i\right) \leq 1 + \theta + \epsilon_i, \forall_i \in [m]. \end{aligned} \tag{7}$$

The final dual form of ssODM is as follows:

$$\min_{\boldsymbol{\mu}\in\mathcal{M}} \max_{\boldsymbol{\alpha}\in\mathcal{A}} \varphi(\boldsymbol{\mu}, \boldsymbol{\alpha}). \tag{8}$$

When a dataset containing a large number of unlabeled samples is input into ssODM, the saddle point $(\widehat{\boldsymbol{\mu}}, \widehat{\boldsymbol{\alpha}})$ of the above problem can be obtained by the random mirror proximal descent method, and the category label of unlabeled samples can be predicted according to $sign(\sum_{k:\,\widehat{y}_k\in\mathcal{B}}\widehat{\mu}_k^*\widehat{\boldsymbol{y}}_k)$.

DYNAMIC STATE TRANSITION ALGORITHM

The expression framework of the dynamic state transition algorithm is as follows:

$$\begin{cases} x_{k+1} = A_k x_k + B_k u_k \\ y_{k+1} = f\left(x_{k+1}\right) \end{cases}, \tag{9}$$

where $x_k = [x_1, x_2, \ldots, x_n]^{\mathrm{T}}$ is the candidate solution of the optimization problem and A_k and B_k are the state transformation operators. u_k represents the control variable, that is, the current and historical state function. $f(\cdot)$ represents the fitness function.

The four transformation operators of the dynamic state transition algorithm covered the fast rotation transformation operator, translation transformation operator, expansion transformation operator, and axesion transformation operator.

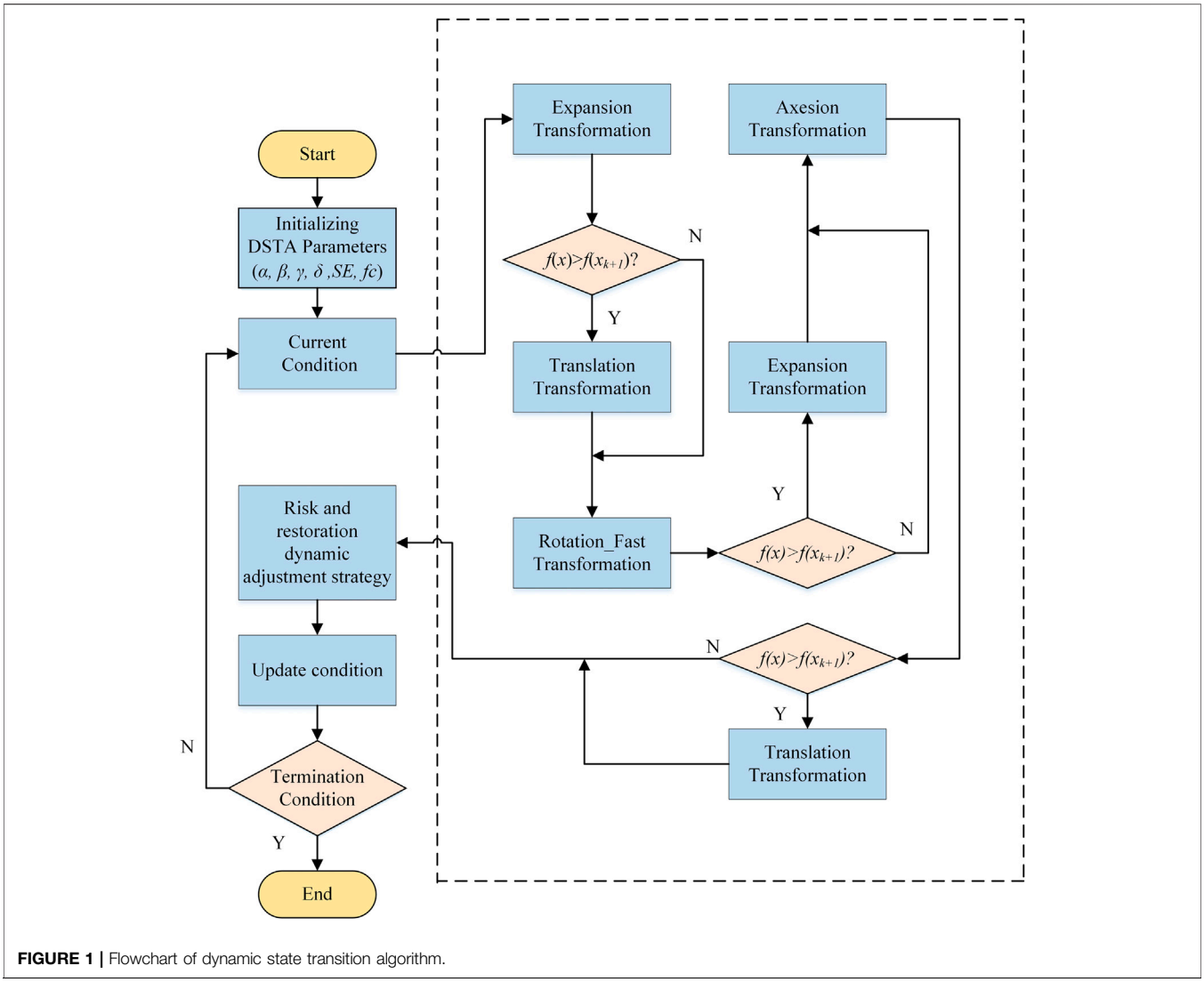

FIGURE 1 | Flowchart of dynamic state transition algorithm.

Fast Rotation Transformation Operator

$$x_{k+1} = x_k + \alpha \widehat{R}_r \frac{u}{u_2}, \tag{10}$$

where α represents the rotation factor and $\widehat{R}_r \in \mathbf{R}^{n\times n}$ is the uniformly distributed random matrix. u represents the random variable that was uniformly distributed on $[-1, 1]$. $\|\cdot\|_2$ is the second norm of the vector. It was found that the fast rotation transformation operator was provided with lower computational complexity, local searchability, and a hypersphere with a radius of α being the search range through the comparison with the rotation change operator.

Translation Transformation Operator

$$x_{k+1} = x_k + \beta R_t \frac{x_k - x_{k-1}}{x_k - x_{k-12}}, \tag{11}$$

where β represents the translation factor. The values of $R_t \in \mathbf{R}$ were uniformly distributed with the range of $[0, 1]$. The translation transformation operator was a heuristic search operator performing the linear search with β being the maximum step length.

Expansion Transformation Operator

$$x_{k+1} = x_k + \gamma R_e x_k, \tag{12}$$

where γ represents the stretching factor. $R_e \in \mathbf{R}^{n\times n}$ is the diagonal matrix whose element value was not equal to zero and was in line with Gaussian distribution. The stretching transformation operator was a global search operator that could stretch all elements in x_k to $(-\infty, +\infty)$ to further search the whole space.

Axesion Transformation Operator

$$x_{k+1} = x_k + \delta R_a x_k, \tag{13}$$

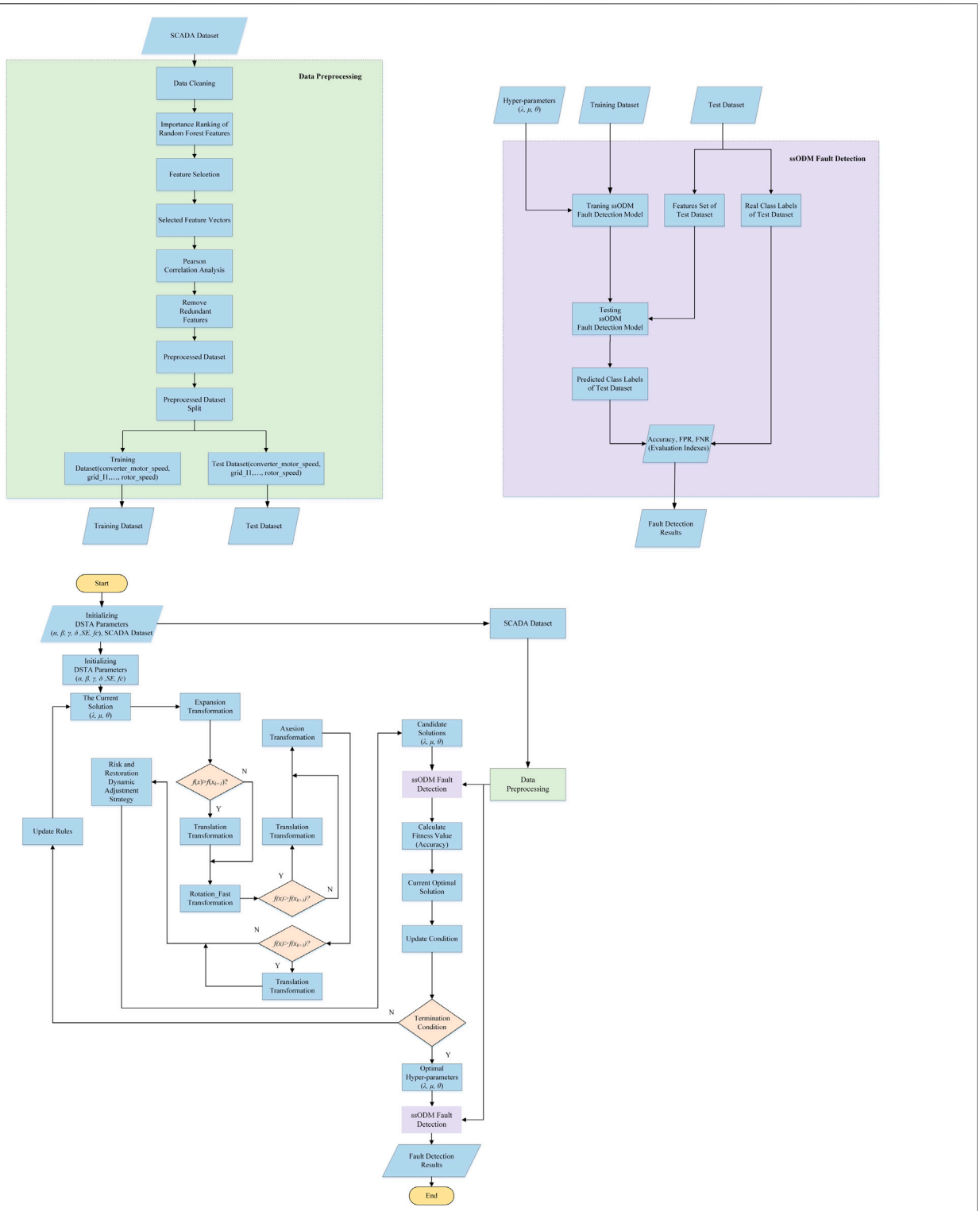

FIGURE 2 | Fault detection flowchart of wind turbine pitch system based on ssODM-DSTA.

`where δ represents the axesion factor. $R_a \in \mathbf{R}^{n\times n}$ is the sparse random diagonal matrix whose value was not equal to zero and was in line with Gaussian distribution. The axesion transformation operator was also a heuristic search operator that could carry out a one-dimensional search along the coordinate axis.

TABLE 1 | Data structures of three faults of pitch system.

Fault type	Number of data samples	Number of data features
Emergency stop fault of pitch system	1,158	58
CAN communication fault of blade 1 servo driver	1,679	58
Low-temperature fault of the blade 2 axle box of the pitch	2,144	58

The flowchart of the DSTA is shown in **Figure 1**.

FAULT DETECTION FOR WIND TURBINE PITCH SYSTEM

Concerning the fault detection process of the wind turbine pitch system, it was roughly made of data screening, data preprocessing, feature selection, data segmentation, training fault detection model and hyperparameter optimization, fault detection model testing, and detection result evaluation. The fault detection flowchart of the ssODM-DSTA wind turbine pitch system is shown **Figure 2**.

Dataset Description

The experimental data originated from the real-time operation data of the SCADA system of 1.5 MW double-fed wind turbines in the domestic wind farm for one year. The data sampling interval was 1 s. The data of the wind turbines in the wind farm in the 12th month were selected and the data of three kinds of wind turbines' pitch faults from half an hour before the start of the faults to half an hour after the end of the faults were intercepted according to the fault codes. Three kinds of pitch system faults are used as fault detection targets: emergency stop fault of the pitch system, CAN communication fault of the blade 1 servo driver, and low-temperature fault of the blade 2 axle box of the pitch, respectively.

A pitch emergency stop fault is a fault triggered when the pitch safety chain acts. The treatment method of this fault is to check whether the pitch safety chain is closed and check the pitch fault specifically.

CAN communication fault of the blade 1 servo driver is the CANBUS communication fault between pitch PLC and pitchmaster of blade 1. The troubleshooting method is to check the wiring between the main control cabinet EL6751 and the X5 terminal on the pitchmaster in the shaft cabinet 2; check whether there is 24V DC voltage between x5-6 and x5-9; check the resistance between x5-2 and x5-7 (60 Ω); check whether the axis 2 servo driver is normal.

The low-temperature fault of the blade 2 axle box of the pitch is a temperature fault of the pitch system. The starting reason is generally the sensor fault. The fault can be eliminated and handled by checking various indicators of the sensor. The three fault data structures are shown in **Table 1**.

Data Cleaning and Preprocessing

The data of the wind turbine pitch system was fed back to the SCADA system in the form of a signal after the information was collected through various sensors such as current, voltage, and speed. The sensors were precision components. They were easily disturbed by the environment and their abnormalities in the monitoring process, which often led to abnormalities and vacancies in the data output of SCADA. The fault detection model had high requirements for the quality of data, so it was necessary to preprocess the data such as standardization, normalization, elimination of outliers, vacancy values, and all "0" eigenvectors (Tang M et al., 2020; Tang S et al., 2020).

Feature selection belongs to a data dimensionality reduction process. The commonly used feature selection methods cover the random forest method (Charvent et al., 2021), extreme gradient boosting (XGBoost)method (Chen et al., 2020), Pearson's correlation analysis method (Jebli et al., 2021), and categorical boosting (CatBoost) method (Yuan et al., 2021). The random forest algorithm was used for ranking the importance of all sub-features of the preprocessed fault data of the pitch system with the wind speed as the target feature. Then, the threshold (mean value of all feature importance scores) was set to screen the features that were highly related to the target variable. On the premise of ensuring no loss of data information content, the number of features of the original data was reduced from 58 to 30, which effectively reduced the difficulty of learning the fault detection model of the wind turbine pitch system. The feature importance ranking of the fault data of the pitch system based on the random forest is shown in **Table 2** (the bold part of No. 1–No. 30 was the 30 features that were saved after screening).

As redundant features with a high correlation may exist between features and the Pearson-related analysis method could be used for ensuring the equal relationship between

TABLE 2 | Feature importance ranking of the fault data of the pitch system based on the random forest.

Serial number	Feature variable name	Random forest feature importance score
1	**main_loop_rotor_speed_demand**	**0.056641**
2	**pitch_position_1**	**0.055856**
30	**converter_power**	**0.009573**
57	pitch_Atech_SG1_error_code_3	0.001598
58	average_wind_speed_30s	0.001518

The bold part of No. 1–30 was the 30 features that were saved after screening.

TABLE 3 | Correlation results between the features.

Pearson correlation coefficient	main_loop_rotor_speed_demand	pitch_position_1	pitch_Atech_position_target_1	pitch_position_target_1
converter_motor_speed	**0.993101**	**−0.9962**	**−0.99644**	**−0.99646**
pitch_Atech_actual_pitch_angle_2	**−0.99292**	**0.999211**	**0.999689**	**0.9997**
average_pitch_position_blade_10m	0.396733	−0.37188	-0.37148	−0.37142

The bold parts show that the feature correlation between different parts of the pitch system was also extremely high and the correlation coefficient between these features was close to "1" with the basically same effect in the data set.

TABLE 4 | Final data structures of 3 faults of pitch system after feature selection.

Fault type	Number of data samples	Number of data features
Emergency stop fault of the pitch system	1,158	24
Can communication fault of the blade 1 servo driver	1,679	24
Low-temperature fault of the blade 2 axle box of the pitch	2,144	24

TABLE 5 | Meaning and value range of ssODM Hyperparameters.

Hyperparameter	Meaning	Value range
λ	Adjust the weight of the margin variance of the objective function	$[2^0, 2^{20}]$
μ	Adjust the weight of sample margin deviation from the positive and negative directions of margin mean	(0.1, 0.9)
θ	Control the sparsity of the model and reduce the number of sample support vectors	(0.1, 0.9)

features and analyzing the correlation degree between features, the Pearson-related analysis method was adopted to analyze the feature correlation of 30 pieces of screened fault data of the pitch system and remove the redundant features with a high correlation for further reducing the data capacity. The correlation results between the features are shown in **Table 3**.

Pearson correlation coefficient was an index describing the intensity of feature correlation with a value range of [-1, 1]. The closer the absolute value of the coefficient was to "1", the stronger the correlation was. The bold parts of **Table 3** show that the feature correlation between different parts of the pitch system was also extremely high and the correlation coefficient between these features was close to "1" with basically the same effect in the dataset. These features that belonged to redundant features were eliminated in the fault dataset of the pitch system and the remaining features were constructed into a new sample dataset. The final data structure is shown in **Table 4**.

Improved Semi-supervised Optimal Margin Distribution Machine

The ssODM was equipped with three hyperparameters λ, μ, and θ representing the margin variance balance hyperparameter, margin deviation balance hyperparameter, and insensitive loss function, respectively. The meaning and value range of ssODM hyperparameters are shown in **Table 5**.

The dynamic state transition algorithm was adopted to optimize the three hyperparameters of ssODM. The classification accuracy of ssODM was taken as the fitness function to determine the update and selection of hyperparameters by the dynamic state transition algorithm. The pseudo-code of the improved ssODM is shown in **Algorithm 1**.

Algorithm 1 Improved semi-supervised optimal margin distribution machine.

$Best \leftarrow Best_0(\lambda_0;\ \mu_0;\ \theta_0)$
repeat
if $\alpha(\beta,\gamma,\delta) < \alpha_{\min}(\beta_{min},\gamma_{min},\delta_{min})$ then
$\alpha(\beta,\gamma,\delta) \leftarrow \alpha_{\max}(\beta_{max},\gamma_{max},\delta_{max})$
end if
$\lambda \leftarrow Best(1)$
$\mu \leftarrow Best(2)$
$\theta \leftarrow Best(3)$
ssODM$\leftarrow(\lambda,\mu,\theta,$ training dataset)
$accuracy$(ssODM)$\leftarrow$testing dataset
$funfcn \leftarrow accuracy$(ssODM)
$[Best, fBest] \leftarrow$rotation_fast$(funfcn, Best, SE, \alpha, \beta)$
$[Best, fBest] \leftarrow$expansion$(funfcn, Best, SE, \beta, \gamma)$
$[Best, fBest] \leftarrow$axesion$(funfcn, Best, SE, \beta, \delta)$
if $fBest < fBest^*$ then
$Best^* \leftarrow Best$
$fBest^* \leftarrow fBest$
end if
if $rand < p_{rest}$ then $\#p_{rest}$: restoration in probability
$Best^* \leftarrow Best$
$fBest^* \leftarrow fBest$
end if
$\alpha(\beta,\gamma,\delta) \leftarrow \frac{\alpha(\beta,\gamma,\delta)}{f_c}$
Until the specified termination criterion is met
Output *Best*

TABLE 6 | Names and meanings of evaluation indexes of the model.

Indicator name	Meaning
TP	Positive samples are predicted to be positive samples
FP	Negative samples are predicted to be positive samples
TN	Positive samples are predicted to be negative samples
FN	Negative samples are predicted to be negative samples

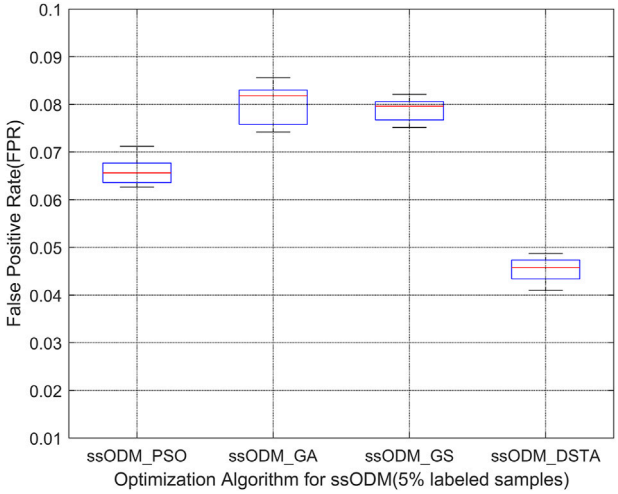

FIGURE 3 | Boxplot of FPR of fault detection regarding the pitch emergency stop (5% labeled samples).

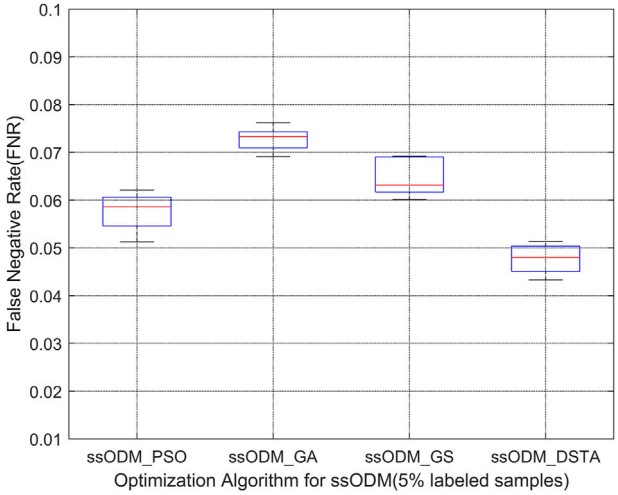

FIGURE 4 | Boxplot of FNR of fault detection regarding the pitch emergency stop (5% labeled samples).

Performance Evaluation Index of Fault Detection Model

The correct prediction of the purity of normal samples and fault samples was deemed to be an index to evaluate the quality of a model. To verify the effectiveness of fault detection of ssODM-DSTA, the FPR and the FNR proposed based on the confusion matrix were taken as the evaluation indexes of the model:

$$\mathrm{FPR} = \frac{FP}{FP + TN}, \tag{14}$$

$$\mathrm{FNR} = \frac{FN}{FN + TP}, \tag{15}$$

where the specific meanings of TP, FP, TN, and FN are shown in **Table 6**.

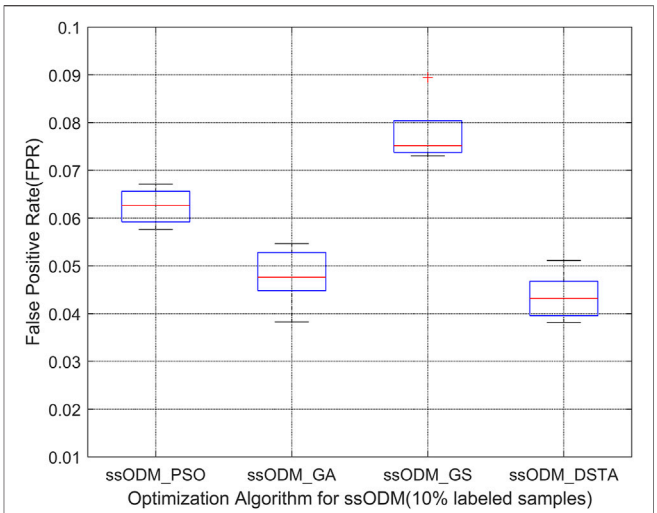

FIGURE 5 | Boxplot of FPR of fault detection regarding the pitch emergency stop (10% labeled samples).

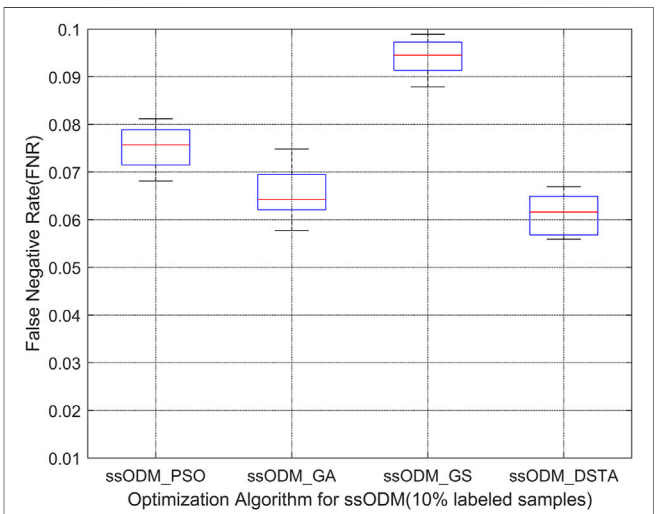

FIGURE 6 | Boxplot of FNR of fault detection regarding the pitch emergency stop (10% labeled samples).

EXPERIMENTAL RESULTS

For verifying the effectiveness of using DSTA to optimize the hyperparameters of ssODM, the PSO, GA, and GS were

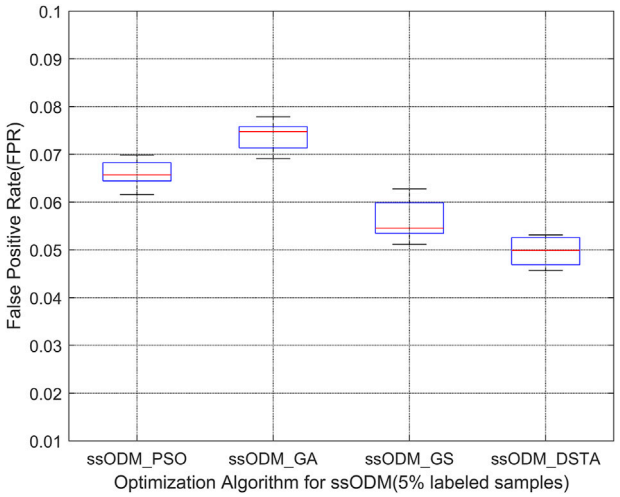

FIGURE 7 | Boxplot of FPR of CAN communication fault detection regarding blade 1 servo driver (5% labeled samples).

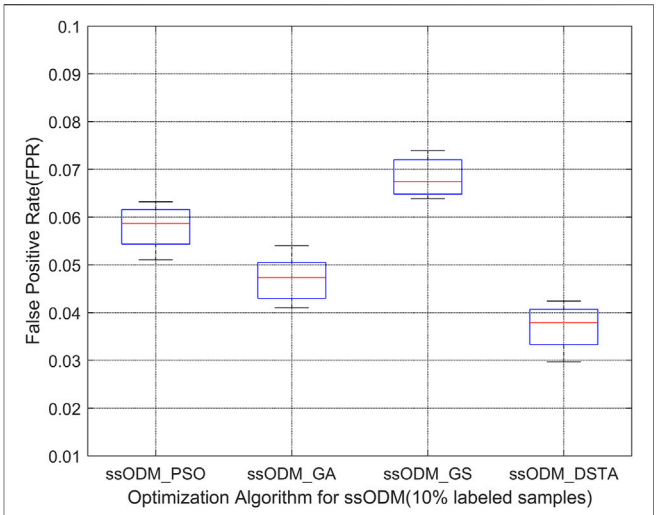

FIGURE 9 | Boxplot of FPR of CAN communication fault detection regarding blade 1 servo driver (10% labeled samples).

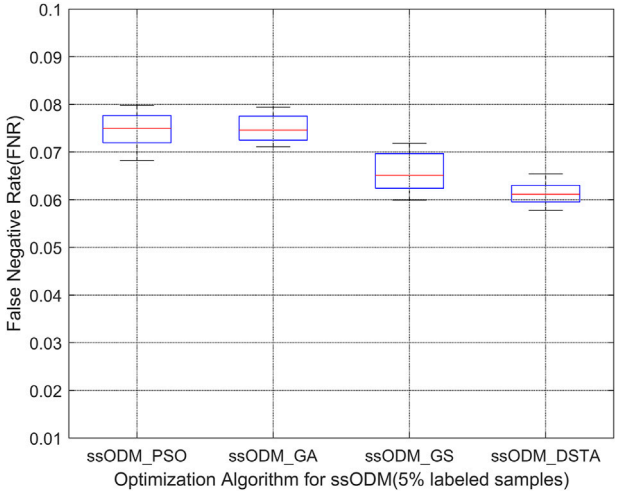

FIGURE 8 | Boxplot of FNR of CAN communication fault detection regarding blade 1 servo driver (5% labeled samples).

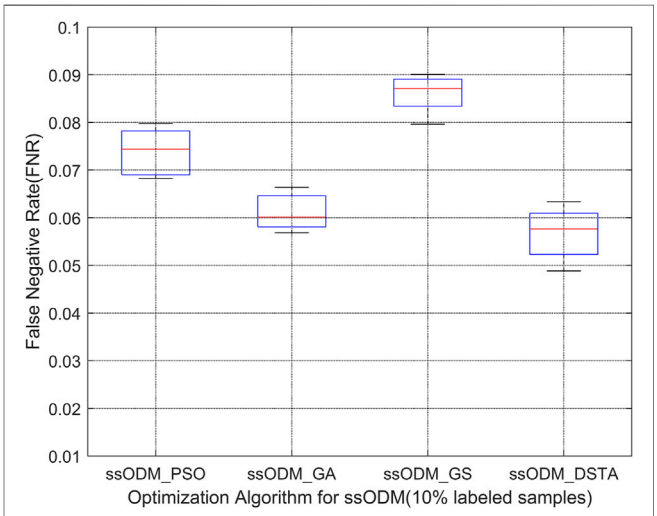

FIGURE 10 | Boxplot of FNR of CAN communication fault detection regarding blade 1 servo driver (10% labeled samples).

introduced to carry out the optimization comparison for the ssODM, respectively. The experimental data were from wind turbine pitch system fault data (the fault data of pitch system are set as 5% labeled data and 10% labeled data for experiments, respectively; the normal sample label is 1 and the fault sample label is—1), as shown in **Table 4**.

When the experimental sample was the emergency stop fault of the pitch, **Figures 3**, **4** represent the FPR of fault detection and the FNR of fault detection, respectively (5% labeled samples). **Figures 5**, **6** represented the FPR of fault detection and the FNR of fault detection (10% labeled samples).

When the experimental sample was the CAN communication fault of blade 1 servo driver, **Figures 7**, **8** represent the FPR of fault detection and the FNR of fault detection, respectively (5% labeled samples). **Figures 9**, **10** represent the FPR of fault detection and the FNR of fault detection (10% labeled samples).

When the experimental sample was the low-temperature fault of the blade 2 axle box of the pitch, **Figures 11**, **12** represent the FPR of fault detection and the FNR of fault detection, respectively (5% labeled samples). **Figures 13**, **14** represent the FPR of fault detection and the FNR of fault detection (10% labeled samples).

The above results have demonstrated that the FPR and the FNR of fault detection regarding the faults for the wind turbine pitch system by ssODM-DSTA were the lowest among the four comparison algorithms. It can be concluded that using DSTA to optimize ssODM can obtain super parameters that more meet the performance requirements of the pitch system fault detection model, effectively improve the

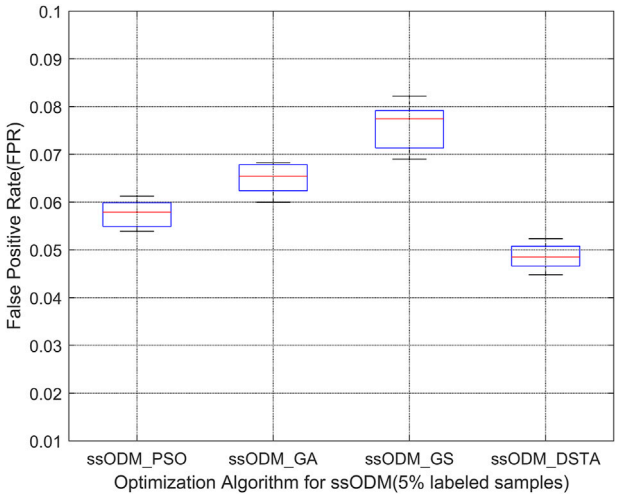

FIGURE 11 | Boxplot of FPR of low-temperature fault detection regarding the blade 2 axle box of the pitch (5% labeled samples).

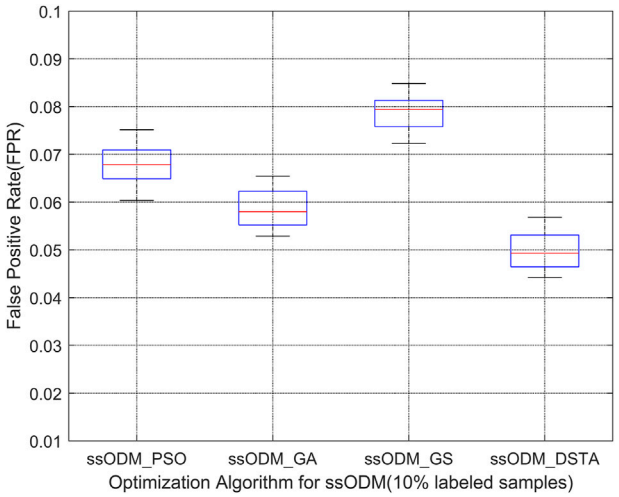

FIGURE 13 | Boxplot of FPR of low-temperature fault detection regarding the blade 2 axle box of the pitch h (10% labeled samples).

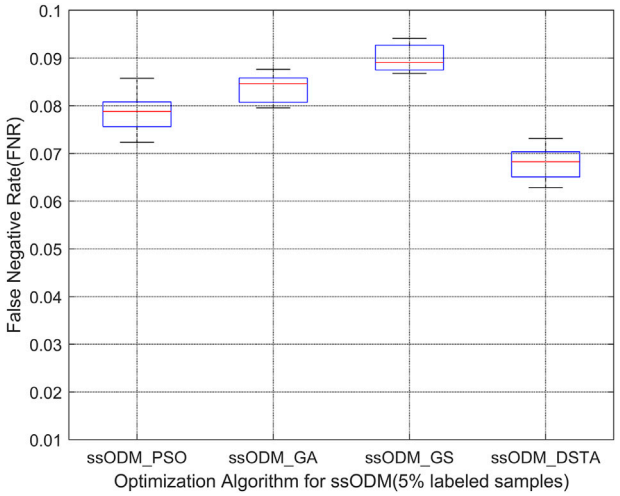

FIGURE 12 | Boxplot of FNR of low-temperature fault detection regarding the blade 2 axle box of the pitch (5% labeled samples).

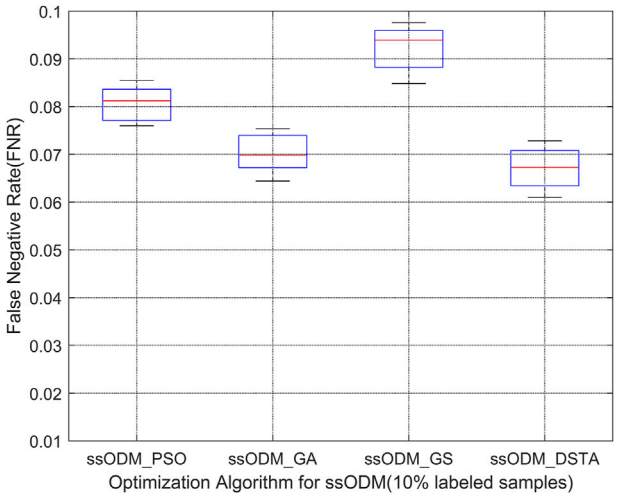

FIGURE 14 | Boxplot of FNR of low-temperature fault detection regarding the blade 2 axle box of the pitch h (10% labeled samples).

classification performance of ssODM, and reduce the error of wind turbine pitch system fault detection.

CONCLUSION

In terms of the problem of obtaining the optimal hyperparameters in the fault detection model of the wind turbine pitch system, the DSTA was used for optimizing the three hyperparameters of ssODM. To verify the effectiveness of this method, ssODM-DSTA was compared with ssODM-PSO, ssODM-GA, and ssODM-GS. The experimental data originated from the three kinds of pitch system fault data preprocessed by the random forest method and Pearson correlation analysis method, including the emergency stop fault data of the pitch system, CAN communication fault data of the blade 1 servo driver, and the low-temperature fault data of the blade 2 axle box of the pitch. The experimental results showed that the ssODM-DSTA had a strong fault detection ability for the fault of the wind turbine pitch system. It was provided with a lower FPR and FNR than those of the model using the other three kinds of parameter optimization algorithms. It was proved that the fault detection method of the wind turbine pitch system based on the ssODM-DSTA had an outstanding performance.

Concerning the wind turbine fault detection based on machine learning, the fault detection model could not be fully learned due to a shortage of labels, seriously affecting the accuracy of fault detection. Consequently, the application of unsupervised learning to wind turbine fault detection could further weaken the weight of

labels and reduce the cost of fault detection when compared with supervised learning and semi-supervised learning.

DATA AVAILABILITY STATEMENT

The data analyzed in this study are subject to the following licenses/restrictions: data used to support the findings of this study are currently under embargo while the research findings are commercialized. Requests to access these datasets should be directed to whw_xy@hbuas.edu.cn.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication.

FUNDING

This work was supported in part by the National Natural Science Foundation of China (Grant Nos. 62173050 and 61403046), the Natural Science Foundation of Hunan Province, China (Grant No. 2019JJ40304), Changsha University of Science and Technology "The Double First Class University Plan" International Cooperation and Development Project in Scientific Research in 2018 (Grant No. 2018IC14), the Research Foundation of the Education Bureau of Hunan Province (Grant No.19K007), Hunan Provincial Department of Transportation 2018 Science and Technology Progress and Innovation Plan Project (Grant No. 201843), Energy Conservation and Emission Reduction Hunan University Student Innovation and Entrepreneurship Education Center, Innovative Team of Key Technologies of Energy Conservation, Emission Reduction and Intelligent Control for Power-Generating Equipment and System, CSUST, Hubei Superior and Distinctive Discipline Group of Mechatronics and Automobiles (XKQ2021003 and XKQ2021010), Major Fund Project of Technical Innovation in Hubei (Grant No. 2017AAA133), and Guangxi Key Laboratory of Trusted Software (No.201728), Graduate Scientific Research Innovation Project of Changsha University of Science & Technology (No. 2021-89).

REFERENCES

Blaabjerg, F., and Ke Ma, K. (2013). Future on Power Electronics for Wind Turbine Systems. *IEEE J. Emerg. Sel. Top. Power Electron.* 1 (3), 139–152. doi:10.1109/JESTPE.2013.2275978

Blaabjerg, F., Liserre, M., and Ma, K. (2012). Power Electronics Converters for Wind Turbine Systems. *IEEE Trans. Ind. Applicat.* 48 (2), 708–719. doi:10.1109/TIA.2011.2181290

Chavent, M., Genuer, R., and Saracco, J. (2021). Combining Clustering of Variables and Feature Selection Using Random Forests. *Commun. Stat. - Simulation Comput.* 50 (2), 426–445. doi:10.1080/03610918.2018.1563145

Chen, C., Zhang, Q., Yu, B., Yu, Z., Lawrence, P. J., Ma, Q., et al. (2020). Improving Protein-Protein Interactions Prediction Accuracy Using XGBoost Feature Selection and Stacked Ensemble Classifier. *Comput. Biol. Med.* 123, 103899. doi:10.1016/j.compbiomed.2020.103899

Chen, X. (2020). Fault Diagnosis of High Power Grid Wind Turbine Based on Particle Swarm Optimization BP Neural Network during COVID-19 Epidemic Period. *Ifs* 39, 9027–9035. doi:10.3233/JIFS-189301

Chong, Y., Ding, Y., Yan, Q., and Pan, S. (2020). Graph-based Semi-supervised Learning: A Review. *Neurocomputing* 408, 216–230. doi:10.1016/j.neucom.2019.12.130

Dai, J., Zheng, J., Pan, H., and Pan, Z. (2017). Rolling Bearing Fault Diagnosis Method Based on Composite Multiscale Entropy and Laplacian SVM. *Zhongguo Jixie Gongcheng/China Mech. Eng.* 28 (11), 1339–1346.

Hu, X., Zhang, H., Ma, D., and Wang, R. (2021). A tnGAN-Based Leak Detection Method for Pipeline Network Considering Incomplete Sensor Data. *IEEE Trans. Instrum. Meas.* 70, 1–10. doi:10.1109/tim.2020.3045843

Jebli, I., Belouadha, F.-Z., Kabbaj, M. I., and Tilioua, A. (2021). Prediction of Solar Energy Guided by pearson Correlation Using Machine Learning. *Energy* 224, 120109. doi:10.1016/j.energy.2021.120109

Li, W. (2010). Gear Incipient Fault Diagnosis Using Graph Theory and Transductive Support Vector Machine. *Jme* 46, 82–88. doi:10.3901/JME.2010.23.082

Liming, S., and Bo, Y. (2020). Nonlinear Robust Fractional-Order Control of Battery/SMES Hybrid Energy Storage Systems. *Power Syst. Prot. Control.* 48 (22), 76–83.

Long, W., Jiao, J., Liang, X., Wu, T., Xu, M., and Cai, S. (2021a). Pinhole-Imaging-Based Learning Butterfly Optimization Algorithm for Global Optimization and Feature Selection. *Appl. Soft Comput.* 103, 107146.

Long, W., Wu, T., Xu, M., Tang, M., and Cai, S. (2021b). Parameters Identification of Photovoltaic Models by Using an Enhanced Adaptive Butterfly Optimization Algorithm. *Energy* 229, 120750.

Malik, H., and Mishra, S. (2017). Artificial Neural Network and Empirical Mode Decomposition Based Imbalance Fault Diagnosis of Wind Turbine Using TurbSim, FAST and Simulink. *IET Renew. Power Generation*, 11(6). Institution of Engineering and Technology, 889–902. doi:10.1049/iet-rpg.2015.0382

Mao, W., Tian, S., Fan, J., Liang, X., and Safian, A. (2020). Online Detection of Bearing Incipient Fault with Semi-supervised Architecture and Deep Feature Representation. *J. Manufacturing Syst.* 55, 179–198. doi:10.1016/j.jmsy.2020.03.005

Odofin, S., Bentley, E., and Aikhuele, D. (2018). Robust Fault Estimation for Wind Turbine Energy via Hybrid Systems. *Renew. Energ.* 120, 289–299. doi:10.1016/j.renene.2017.12.031

Qian, W., and Wang, J. (2020). An Improved Seasonal GM(1,1) Model Based on the HP Filter for Forecasting Wind Power Generation in China. *Energy* 209, 118499. doi:10.1016/j.energy.2020.118499

Shen, Z., Chen, X., Zhang, X., and He, Z. (2012). A Novel Intelligent Gear Fault Diagnosis Model Based on EMD and Multi-Class TSVM. *Measurement* 45, 30–40. doi:10.1016/j.measurement.2011.10.008

Song, D., Chang, Q., Zheng, S., Yang, S., Yang, J., and Hoon Joo, Y. (2021). Adaptive Model Predictive Control for Yaw System of Variable-Speed Wind Turbines. *J. Mod. Power Syst. Clean Energ.* 9 (1), 219–224. doi:10.35833/MPCE.2019.000467

Sun, Q., Han, R., Zhang, H., Zhou, J., and Guerrero, J. M. (2015). A Multiagent-Based Consensus Algorithm for Distributed Coordinated Control of Distributed Generators in the Energy Internet. *IEEE Trans. Smart Grid* 6 (6), 3006–3019. doi:10.1109/TSG.2015.2412779

Tan, Z. H., Tan, P., Jiang, Y., and Zhou, Z. H. (2020). Multi-label Optimal Margin Distribution Machine. *Mach Learn.* 109 (3), 623–642. doi:10.1007/s10994-019-05837-8

Tang, M., Chen, Y., Wu, H., Zhao, Q., Long, W., Sheng, V. S., et al. (2021). Cost-Sensitive Extremely Randomized Trees Algorithm for Online Fault Detection of Wind Turbine Generators. *Front. Energ. Res.* 9 (234). doi:10.3389/fenrg.2021.686616

Tang, M., Ding, S. X., Yang, C., Cheng, F., Shardt, Y. A. W., Long, W., et al. (2019). Cost-sensitive Large Margin Distribution Machine for Fault Detection of Wind Turbines. *Cluster Comput.* 22 (3), 7525–7537. doi:10.1007/s10586-018-1854-3

Tang, M., Zhao, Q., Ding, S. X., Wu, H., Li, L., Long, W., et al. (2020). An Improved LightGBM Algorithm for Online Fault Detection of Wind Turbine Gearboxes. *Energies* 13, 807. doi:10.3390/en13040807

Tang, S., Yuan, S., and Zhu, Y. (2020). Data Preprocessing Techniques in Convolutional Neural Network Based on Fault Diagnosis towards Rotating Machinery. *IEEE Access* 8, 149487–149496. doi:10.1109/ACCESS.2020.3012182

Tuerxun, W., Chang, X., Hongyu, G., Zhijie, J., and Huajian, Z. (2021). Fault Diagnosis of Wind Turbines Based on a Support Vector Machine Optimized by the Sparrow Search Algorithm. *IEEE Access* 9, 69307–69315. doi:10.1109/ACCESS.2021.3075547

Wang, R., Sun, Q., Hu, W., Li, Y., Ma, D., and Wang, P. (2021). SoC-Based Droop Coefficients Stability Region Analysis of the Battery for Stand-Alone Supply Systems with Constant Power Loads. *IEEE Trans. Power Electron.* 36 (7), 7866–7879. doi:10.1109/TPEL.2021.3049241

Xi, L., Wu, J., Xu, Y., and Sun, H. (2021). Automatic Generation Control Based on Multiple Neural Networks With Actor-Critic Strategy. *IEEE Trans. Neural Netw. Learn. Syst.* 32 (6), 2483–2493. doi:10.1109/TNNLS.2020.3006080

Yang, J., Fang, L., Song, D., Su, M., Yang, X., Huang, L., et al. (2021). Review of Control Strategy of Large Horizontal-axis Wind Turbines Yaw System. *Wind Energy* 24 (2), 97–115. doi:10.1002/we.2564

Yao, L., Fang, Z., Xiao, Y., Hou, J., and Fu, Z. (2021). An Intelligent Fault Diagnosis Method for Lithium Battery Systems Based on Grid Search Support Vector Machine. *Energy* 214, 118866. doi:10.1016/j.energy.2020.118866

Yuan, Z., Zhou, T., Liu, J., Zhang, C., and Liu, Y. (2021). Fault Diagnosis Approach for Rotating Machinery Based on Feature Importance Ranking and Selection. *Shock and Vibration* 2021, 1–17. doi:10.1155/2021/8899188

Zhang, L., Ji, S., Gu, S., Huang, X., Palmer, J. E., Giewont, W., et al. (2021). Design Considerations for High-Voltage Insulated Gate Drive Power Supply for 10-kV SiC MOSFET Applied in Medium-Voltage Converter. *IEEE Trans. Ind. Electron.* 68 (7), 5712–5724. doi:10.1109/TIE.2020.3000131

Zhang, L., and Ruan, X. (2019). Control Schemes for Reducing Second Harmonic Current in Two-Stage Single-phase Converter: An Overview from DC-Bus Port-Impedance Characteristics. *IEEE Trans. Power Electron.* 34 (10), 10341–10358. doi:10.1109/TPEL.2019.2894647

Zhang, T., and Zhou, Z.. (2018). "Semi-Supervised Optimal Margin Distribution Machines," in Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence; 2018; Stockholm, Sweden; 3104–3110. doi:10.24963/ijcai.2018/431

Zhang, T., and Zhou, Z.-H. (2014). "Large Margin Distribution Machine," in Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining, New York, USA (New York, NY, United States: Association for Computing Machinery), 313–322. doi:10.1145/2623330.2623710

Zhang, X., Han, P., Xu, L., Zhang, F., Wang, Y., and Gao, L. (2020). Research on Bearing Fault Diagnosis of Wind Turbine Gearbox Based on 1DCNN-PSO-SVM. *IEEE Access* 8, 192248–192258. doi:10.1109/ACCESS.2020.3032719

Zhang, Y., and Sheng, R. (2021). Fault Diagnosis Method of Mine Motor Based on Support Vector Machine. *Eng* 14 (3), 508–514. doi:10.2174/1872212113666191121122720

Zhang, Y., Zheng, H., Liu, J., Zhao, J., and Sun, P. (2018). An Anomaly Identification Model for Wind Turbine State Parameters. *J. Clean. Prod.* 195, 1214–1227. doi:10.1016/j.jclepro.2018.05.126

Zhang, Z.-Y., and Wang, K.-S. (2014). Wind Turbine Fault Detection Based on SCADA Data Analysis Using ANN. *Adv. Manuf.* 2 (1), 70–78. doi:10.1007/s40436-014-0061-6

Zhao, H., Gao, Y., Liu, H., and Li, L. (2019). Fault Diagnosis of Wind Turbine Bearing Based on Stochastic Subspace Identification and Multi-Kernel Support Vector Machine. *J. Mod. Power Syst. Clean. Energ.* 7 (2), 350–356. doi:10.1007/s40565-018-0402-8

Zhou, X., Shi, P., Lim, C.-C., Yang, C., and Gui, W. (2018). A Dynamic State Transition Algorithm with Application to Sensor Network Localization. *Neurocomputing* 273 (17), 237–250. doi:10.1016/j.neucom.2017.08.010

Zhou, X., Yang, C., Yang, C., and Gui, W. (2012). State Transition Algorithm. *J. Ind. Manage. Optimization* 8 (4), 1039–1056. doi:10.3934/jimo.2012.8.1039

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Energy Research

REVIEW
published: 15 November 2021
doi: 10.3389/fenrg.2021.751066

Review and Perspectives of Machine Learning Methods for Wind Turbine Fault Diagnosis

Mingzhu Tang[1†], Qi Zhao[1], Huawei Wu[2], Ziming Wang[3†], Caihua Meng[1] and Yifan Wang[1]*

[1]School of Energy and Power Engineering, Changsha University of Science & Technology, Changsha, China, [2]Hubei Key Laboratory of Power System Design and Test for Electrical Vehicle, Hubei University of Arts and Science, Xiangyang, China, [3]School of Computer Science and Information Security, Guilin University of Electronic Technology, Guilin, China

Wind turbines (WTs) generally comprise several complex and interconnected systems, such as hub, converter, gearbox, generator, yaw system, pitch system, hydraulic system control system,integration control system, and auxiliary system. Moreover, fault diagnosis plays an important role in ensuring WT safety. In the past decades, machine learning (ML) has showed a powerful capability in fault detection and diagnosis of WTs, thereby remarkably reducing equipment downtime and minimizing financial losses. This study provides a comprehensive review of recent studies on ML methods and techniques for WT fault diagnosis. These studies are classified as supervised, unsupervised, and semi-supervised learning methods. Existing state-of-the-art methods are analyzed and characteristics are discussed. Perspectives on challenges and further directions are also provided.

Keywords: wind turbines, fault diagnosis, supervised learning, unsupervised learning, semi-supervised learning

OPEN ACCESS

Edited by:
Lei Xi,
China Three Gorges University, China

Reviewed by:
Lu Qian,
Wuhan University of Technology, China
Xiaojun Zhou,
Central South University, China

****Correspondence:***
Huawei Wu
whw_xy@hbuas.edu.cn

[†]These authors have contributed equally to this work and share first authorship

Specialty section:
This article was submitted to Smart Grids, a section of the journal Frontiers in Energy Research

Received: *31 July 2021*
Accepted: *17 September 2021*
Published: *15 November 2021*

Citation:
Tang M, Zhao Q, Wu H, Wang Z, Meng C and Wang Y (2021) Review and Perspectives of Machine Learning Methods for Wind Turbine Fault Diagnosis. Front. Energy Res. 9:751066. doi: 10.3389/fenrg.2021.751066

INTRODUCTION

Wind power has gained remarkable attention in the past decade because wind energy is one of the rapidly clean energy sources and has received worldwide support for renewable energy development (MUA, 2017). In recent years, in order to achieve the goal of carbon peak and carbon neutralization, China has commercialized the use of renewable energy, expanded the use of renewable energy, and demonstrated its determination to reach the peak of carbon dioxide emission by 2030 and carbon neutralization by 2060. As the main force of global renewable energy development, China attaches great importance to new energy, especially wind power generation. According to the statistics of the Global Wind Energy Commission (GWEC), the newly installed capacity of the country has reached 65.1 GW 2) in 2019 (Elizondo et al., 2019). The large-scale development and utilization of wind energy have brought huge opportunities for the development of the market economy, and also raised important crucial challenges related to reliability, cost-effectiveness, and energy blade images of the security. On the one hand, wind turbines (WTs) are often located in remote areas, operated in harsh working environments for a long time, and have withstood randomly varying weather conditions, wind shear, temperature, wind speed, and load, thereby frequent WT failures. As shown in **Figure 1**, the highest proportion of fault rate of WT components is the electrical system (Hahn et al., 2007), followed by the control system and sensor. On the other hand, the high cost of operation and maintenance (OM) of WTs underscores the urgency of fault diagnosis. Evidently, fault diagnosis and the timely maintenance of WTs can reduce huge financial losses.

Given the preceding reasons, fault warning and fault diagnosis of WTs should be performed. The fault diagnosis method based on machine learning (ML) is suggested to detect the operating

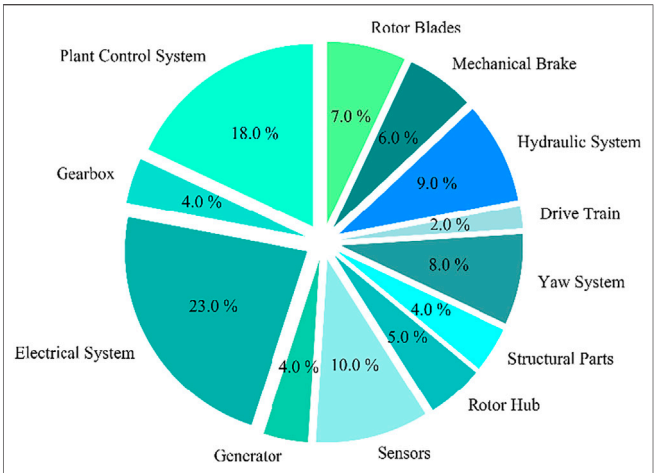

FIGURE 1 | Fault rate of wind turbine components (Hahn et al., 2007).

conditions of the WT for it can minimize the downtime and reduce OM cost WTs, and extend the service life of these turbines. With the advent of the era of fault diagnosis technology, many local and international experts and scholars have proposed some efficient fault diagnosis methods for various components (Liu et al., 2015), such as power system (Qiao and Lu, 2015; Zappalá et al., 2019), mechanical (Wang et al., 2016a; Chen et al., 2016b; Garg and Dahiya, 2017; Salameh et al., 2018), and driving faults (Nasiri et al., 2015; Zeng et al., 2015), etc. Among these methods, generator (Hossain et al., 2015; Yang et al., 2017) and gearbox faults (Wang et al., 2016b; Igba et al., 2016; Teng et al., 2016; Wang et al., 2019) are mostly studied. Fault diagnosis methods are classified into fault diagnosis methods based on analytical models, knowledge-based methods, and data-driven fault diagnosis methods (Chen et al., 2016a).

The analytical model-based WT fault diagnosis methods need to analyze and model the system to achieve real-time diagnoses of the faults, which are often directly related to WT model parameters (Gao et al., 2015; Zhong et al., 2018). With a further understanding of the fault diagnosis mechanism of WT, modeling is implemented to increase the accuracy of fault diagnosis. However, in the process of analytical model-based WT fault diagnosis methods by uses system residuals to model the internal subsystem of the WTs for state estimation and online approximation; nevertheless, this process has difficulty in ensuring the accuracy of fault diagnosis (Liu et al., 2017; Cho et al., 2018). Consequently, inevitable errors and unknown interference terms will result, and the aforementioned process is insufficient to guarantee robustness.

Knowledge-based WT fault diagnosis methods rely on expert experience in wind power-related fields (da Silva et al., 2012; Yang et al., 2016). The accuracy of fault diagnosis results depends on the extensiveness of expert experience and knowledge the level of WT fault diagnosis experts, which lack self-learning and recognition abilities. Knowledge-based WT fault diagnosis methods cannot acquire new knowledge from the diagnosed engineering examples during the operation of WT. Hence, poor diagnosis accuracy may be resulted.

Without relying on prior experiences, data-driven WT fault diagnosis methods uses data mining technology to obtain hidden useful information to characterize the fault and normal states of the system, and eventually realize real-time fault diagnosis (Ding, 2012; Qin, 2012). The WT supervisory control and data acquisition (SCADA) system contains real-time online data and extensive offline data. The use and analysis of data mining is necessary to obtain detailed fault characteristics, thereby realizing real-time WT fault diagnosis. Data-driven WT fault diagnosis methods include the ML, multivariate statistical analysis, signal analysis, and information fusion methods (Yin et al., 2014).

As shown in **Figure 2**, the fault diagnosis methods of WT based on ML can be generally divided into supervised, unsupervised, and semi-supervised learning methods. Although some literature reviews on WT fault diagnosis (Liu et al., 2015) and condition monitoring (de Azevedo et al., 2016) have been published, there still lack of comprehensive review on the ML-based fault diagnosis method of ML. Therefore, the current study provides a systemic and pertinent state-of-the-art review on recent studies on ML methods and techniques that have been used for WT fault diagnosis. In particular, this research summarizes the research methods in WT fault diagnosis, presents the strengths and shortcomings of existing methods, and reveals the challenges and recommendations of future research direction in this domain.

Fault Diagnosis of Wind Turbine

Numerous countries have earlier previously conducted research on WT technology, and European countries and the US have made some progress in fault diagnosis and prediction (Habibi et al., 2019). For example, Siemens' SCADA system is widely used in major wind power generation industries (Dao et al., 2018).

Compared with European and American countries, China's wind power industry started late, but WT fault diagnosis research has made some progress in recent years. Since the progress and development of artificial intelligence and ML in recent years, the

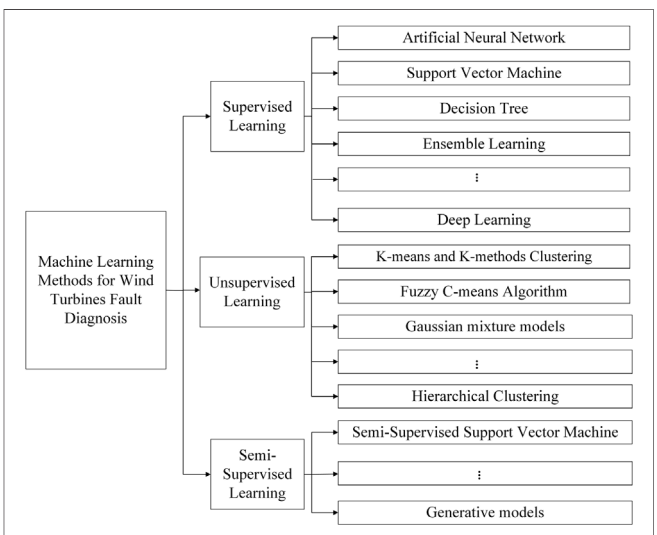

FIGURE 2 | ML methods for WT fault diagnosis.

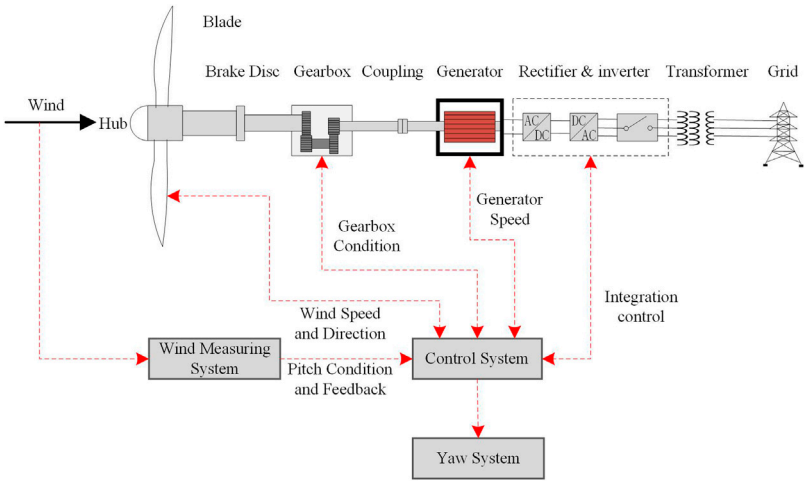

FIGURE 3 | Main structure of WT.

fault diagnosis methods of WT have been intensively studied. The WT structure is shown in **Figure 3**. The main components of WT include wind wheel, gearbox, generator, converter, yaw system, pitch system, hydraulic system control system, integration control system, and auxiliary system (Lin et al., 2016).

The wind wheel is key to the energy conversion of WT, and operational stability directly affects the efficiency and safety of WT. As the operating time of WT increases, the failure rate of the wind wheel and other components also increases, which seriously affects the working performance of WT. In a non-stationary state, the frequency component of the WT failure at the generator output will expand over the bandwidth proportional to the speed, thereby making its diagnosis capabilities considerably complicated. Therefore, Dahiya (2018) proposed a fault diagnosis method of WT based on wavelet analysis, using electrical signals to diagnose rotor eccentric faults. The effectiveness of this method under varying speed and load conditions has been verified through experiments.

Gearboxes are one of the important WT, but the most expensive WT sub-assemblies. Gearboxes are often operating under extreme temperature and high speed of rotation, which will cause a high fault rate and irreversible damage to WT. At present, many studies and research have been conducted on the fault diagnosis of WT gearboxes (Salameh et al., 2018). Du et al. (2015) proposed a convex optimization-based WT generator gearbox fault diagnosis method. This method considers identifying multiple features from the superimposed signal of WT gearbox, and makes full use of the potential a priori information, that is, multiple faults with similar spectrum have different morphological waveforms, which can be sparse represented on the joint of redundant dictionaries. The proposed framework is verified by diagnosing multiple faults of gearbox in wind farm. (Zhang et al., 2017) used the Morlet wavelet-based continuous wavelet transform for actual wind turbine gear fault diagnosis. This diagnosis uses the signal analysis method, which has considerably refined time frequency characteristics and achieved satisfactory results.

A generator is the core equipment for generating electricity through WT, which converts kinetic energy into electrical energy. Generators will also experience a high failure rate owing to the harsh environment, large load fluctuation, and diverse operating parameters of this equipment. Numerous publications have specially reviewed the WT generator fault diagnosis, including those involved in avoiding incorrect internal pattern recognition caused by heavy noise, Chen et al. (2016b) extract inherent modulation information by decomposing the signal into mono-components on an orthogonal basis using empirical wavelet transform (EWT). Moreover, before EWT, they applied wavelet spatial adjacent coefficient denoising with data-driven threshold to improve signal-to-noise ratio (SNR), which is considered to be a powerful tool for WT generator fault diagnosis. Yang et al. (2017) considered the shortcomings of sparse representation results affected by dictionaries, and proposed a novel data-driven fault diagnosis method based on shift-invariant dictionary learning and sparse representation for WT generator, which can effectively identify the WT generator. The coefficients obtained can be considerably sparse based on the learned shift-invariant dictionary, and the impulse signal extracted nearly approximating to the real signal.

The converter is a critical component of the WT energy conversion, and the WT outputs current with stable frequency and amplitude to the grid through the converter. Converters have poor stability and are often impacted by high-temperature and high-pressure working circumstances, and the long-term operation will cause irreversible damage to the WT system. Toubakh et al. (Toubakh and Sayed-Mouchaweh, 2016) analyzed the converter fault caused by parameter drift, and proposed a fault diagnosis method of the WT converter based on a hybrid dynamic classifier, which can monitor the normal operation of converters in the discrete mode affected by parameter failure. The parameter drift under conditions is used for fault diagnosis in the early period of the WT converter. Liang et al. (2020) proposed a fault diagnosis method based on WT converters. A series of inherent mode

functions are obtained through the overall empirical mode decomposition processing of the measured output voltage. Thereafter, the standard entropy is calculated according to the inherent model functions statistical characteristics, the extracted information is used to describe the diagnostic characteristics, and the fault diagnosis of the fan system is performed. The diagnostic accuracy is 99.57%, and its performance was impressive.

The yaw system is an important WT component and can drive the WT engine room to revolve around the tower centerline, thereby maintaining the verticality of the wind wheel scanning surface and wind direction vertical. Yaw system failures often occur owing to its harsh operating environment and load fluctuation, thereby affecting the power generation efficiency of WTs. To qualitatively evaluate the zero-offset error of the yaw system, Pei et al. (2018) proposed a data-driven method for WT yaw system fault diagnosis, which can detect the zero-point shifting fault by analyzing the power characteristics of different yaw angles. If the yaw angle measurement error is greater than a predetermined threshold, then the zero-point shift fault will be triggered, which can detect the fault in time and improve the WT performance. In the case of yaw system faults, Ouanas et al. (2018) proposed a fault diagnosis method of WT yaw system based on the signal analysis method. By filtering the inverter signal provided by the yaw drive, the discrete wavelet transform and empirical mode decomposition method were used to eliminate redundant information. Faults from the envelope of the Hilbert transform are detected, thereby verifying its effectiveness.

Pitch control system is the speed control device of WT and can adjust power change by changing the blade angle of attack. Given to the variable external wind conditions of WT and complicated internal system structure of the pitch system, abnormal output power, blade damages of the s, and even unit collapse can easily be caused, in which the failure rate is high. Many studies have proposed fault diagnosis methods for pitch systems. Habibi et al. (2017) proposed the fault diagnosis method of the WT pitch system by using a nonlinear model and presented the problem of maximizing energy extraction by designing the optimal desired state. Experiments have been performed to verify the practicability of the proposed method. Lan et al. (2018) conducted a study based on the adaptive step-by-step sliding window observer's state estimation and fault indicator functions of a pitch system, which can effectively deal with the nonlinear fault distribution function and identify the pitch fault of WTs.

A hydraulic system is an important WT component and plays a essential role in the yaw, pitch, and transmission chain braking of WTs. Hydraulic system function in all-weather, open-air and high-altitude running conditions, which is prone to failures such as oil leakage and spool jamming, thereby making maintenance difficult. In the case of WT hydraulic system faults, Yang et al. (2011) proposed a fault detection method for WT hydraulic system based on the Petri net model. First, Petri net theory is used to establish a model for each discrete operating state of the WT hydraulic pitch system, and a fault Petri net model is built. Thereafter, a system reliability index is obtained based on the fault qualitative analysis and calculation of the Petri net. The Petri net model calculation is simple, which is ultilized to the WT hydraulic system fault diagnosis and has a broad application prospect.

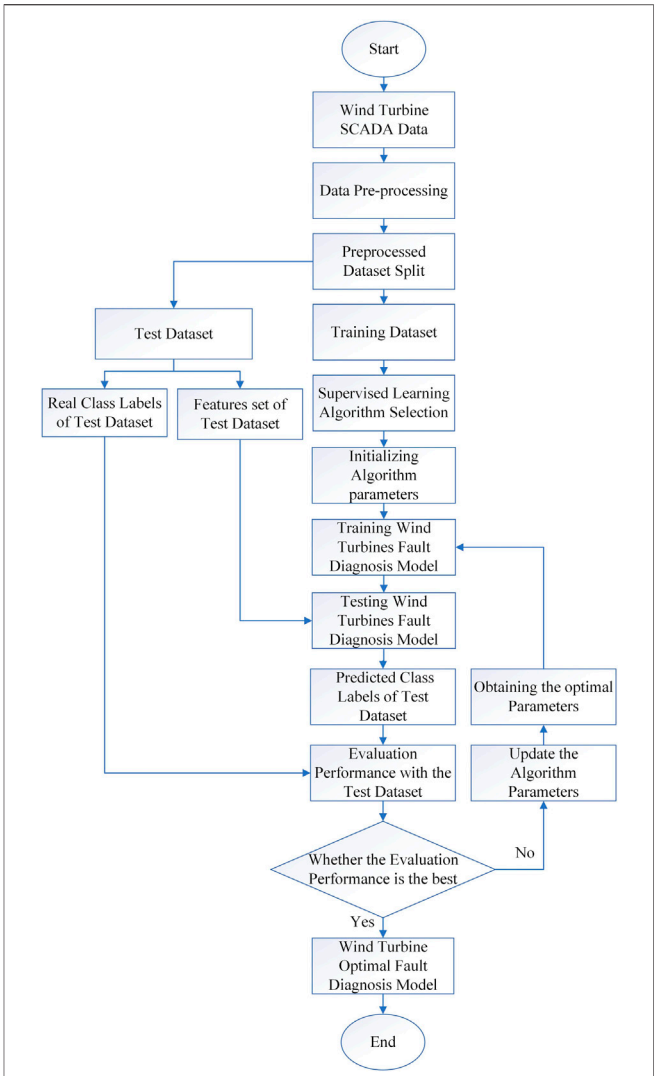

FIGURE 4 | Flowchart of WT fault diagnosis based on the supervised Learning method.

Machine Learning Methods for Wind Turbine Fault Diagnosis

ML refers to a computer that learns from a limited amount of data without specialist intervention to train an inductive model and uses this model thereafter to guide future decisions (Clifton et al., 2013; Stetco et al., 2019). The ML method has been used for fault diagnosis in WT (Leahy et al., 2016), which consists of inputs, outputs, models, and objective functions. Given the WT data sample data $x = \{x_1, x_2, \ldots x_n\}$ (x represents a data set containing n samples) and fault category y, n represent the total number of data sample. Thereafter, we use the training sample $\{x_i, y_i\}_1^M$ ($\{x_i, y_i\}_1^M \in \{x, y\}$) to train the model and obtain the approximate value $f(x)$ to fit the real value y. Moreover, y^* represents the mapping relationship between x and y, and M is the total number of training samples.

$$y^* = argmin_f\left(E_{xy}L(y, f(x))\right) = argmin_f\left(E_x\left(E_y\left(L(y, f(x))\right)|x\right)\right) \quad (1)$$

In **Equation 1**, L represents a loss function, and the average loss of the training set is called empirical risk. The goal of ML is to minimize empirical risk. Frequently employed loss functions include 0–1, square, absolute value, and log loss function.

The problem of overfitting is one of the key issues in the ML method. Therefore, empirical and structural risks should be minimized. The regular term $J(f)$ is introduced to measure the model complexity. The frequently employed regular terms are Lasso and Ridge regression. The final optimized objective function can be expressed as follows:

$$Obj = min\frac{1}{M}\sum_{i=1}^{M}L(y, f(x)) + \lambda J(f) \quad (2)$$

ML methods are divided into supervised, unsupervised, and semi-supervised learning methods (Lei et al., 2020). The current study also classifies the ML-based WT fault diagnosis methods as the supervised, unsupervised, and semi-supervised learning methods, which are analyzed and discussed in the following sections.

Supervised Learning Methods for Wind Turbine Fault Diagnosis

Supervised learning is a process of adjusting classifier parameters using samples of a known class to achieve the desired performance. In supervised learning (Schwenker and Trentin, 2014; Zhou, 2018), the computer is received the example inputs and its required outputs, given input and output, and the target is to learn a general rules of mapping input to output. Supervised learning methods are widely used in the WT fault diagnosis field. As shown in **Figure 4**, supervised learning methods have different algorithms for specific problems. First, we take WT fault diagnosis (Jiménez et al., 2019) as the research object to obtain data from the SCADA of WTs; divide the training, validation, and the test sets and perform data preprocessing on the data set; and normalize the data after processing the missing values. Second, an ML algorithm is chosen to train the training set which is used for modeling. Thirdly, the test set is used to evaluate the model quality. Lastly, an accurate fault classification is obtained by continuously optimizing the fault diagnosis model of the WT.

Artificial Neural Network

Artificial neural network (ANN) (Agatonovic-Kustrin et al., 2000; Xi et al., 2020) is one of the most frequently used supervised learning algorithms. ANN consists of numerous neurons and is divided into input layer, hidden layer, and output layer. ANN is widely used in the fault diagnosis field (Samanta et al., 2003; Saravanan and Ramachandran, 2010). By learning from known fault samples, the mapping relationship between fault characteristics and fault categories is established to detect whether a device is faulty. **Figure 5** shows a three-layer simple WT fault diagnosis model based on ANN, $x_1, x_2, \ldots, x_n$ are the input characteristics of WT, n is the total sample of input characteristics, and m is the total fault types of WT.

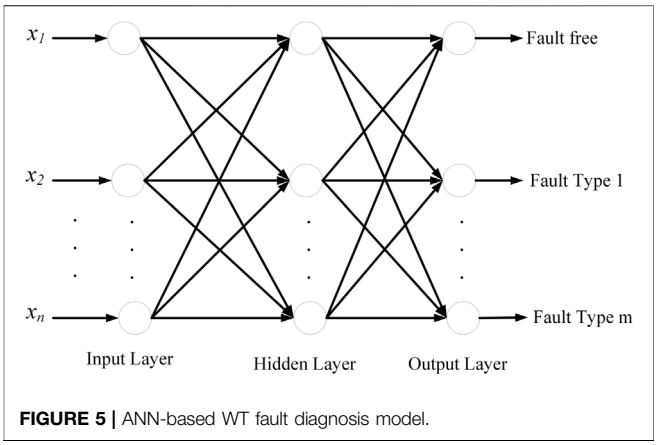

FIGURE 5 | ANN-based WT fault diagnosis model.

The frequently employed neural network methods include adaptive resonance theory (ART), self-organizing map (SOM) neural network and radial basis function (RBF) neural network.

Zhang et al. (Zhang and Wang, 2014). proposed an ANN-based fault diagnosis method for the WT main bearing based on the WT SCADA system data. The difference between the theoretical and the actual parameter values can identify the early stage of the main bearing faults of WT. To decrease the time of ANN for WT fault detection, Bielecki et al. (2014) proposed a hybrid method of ART and RBF neural networks for online detection of the operating status of WT, which can monitor the status of WT in time, identify the early fault conditions and have good real-time performance. However, the actual engineering in a wind farm cannot collect all information on the fault, and the ANN cannot make accurate fault diagnosis. Therefore, Zhao (Zhao et al., 2015) proposed to apply the SOM neural network to the fault diagnosis of WT and to train the network through the sample data of the normal WT state. This is judged whether the wind turbine malfunctions according to the position of the output neuron in the output layer. Accordingly, the SOM neural network method can effectively diagnose the WT fault with good robustness.

Although the fault diagnosis of ANN has high precision and good robustness, this method requires numerous parameters for modeling, and the training model takes a long time. China's wind power industry started late, but WT fault diagnosis research has been developed in the recent years. However, WT fault data samples are considerably lacking, and the accuracy and completeness of the WT data samples directly affect the accuracy of fault diagnosis classification. This issue is currently the main drawback restricting the development of ANN in WT fault diagnosis.

Support Vector Machine

Support vector machine (SVM) is a kernel-based ML method used in regression problems and classification tasks introduced by Vapnik (2013). The main idea is to find two parallel hyperplanes to separate two sets of data in a multi-dimensional space and maximize the margin between the hyperplanes. SVM formulation

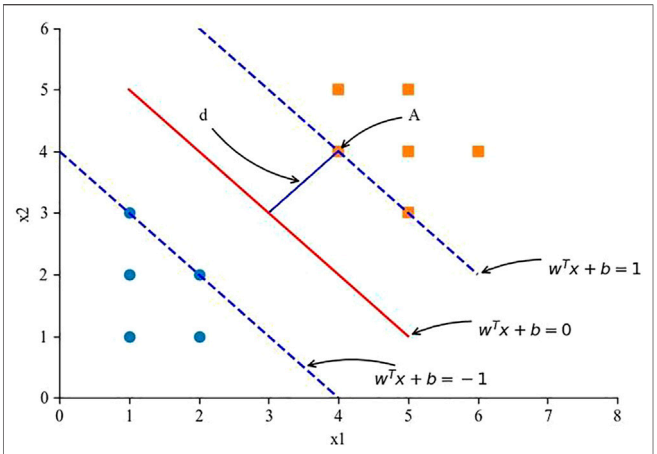

FIGURE 6 | Interval and support vector machine.

ensures that the decision hyperplane is constructed with structural risk minimization to obtain a balance between empirical risk and complexity of model (Deka, 2014). SVM is mainly used in nonlinear problems, by building a classification hyperplane as a decision plane, in which the isolation boundry between negative and positive samples is maximized. As shown in **Figure 6**, any hyperplane can be represented by a normal vector W and a constant b (intercept) as follows:

$$w^T x + b = 1 \tag{3}$$

$$w^T x + b = 0 \tag{4}$$

$$w^T x + b = -1 \tag{5}$$

For point $A(x_1, y_1)$, any two hyperplanes have a geometric interval d.

SVM is to find a hyperplane to make the data points separable, in which the minimum geometric distance is the largest. The SVM solution process can be regarded as the solution process of a convex quadratic problem, which has a global optimal solution. Thus, SVM is widely used in the fault diagnosis field.

To solve the local optimal phenomenon caused by the improper selection of sample parameters, Laouti et al. (2011) chose a radial basis function as the kernel parameter of SVM, which can immediately detect the WT blade pitch positionand generator failureand has good generalization performance. To further solve the problem of overfitting or underfitting caused by the improper selection of nuclear parameters, Tang et al. (2014) proposed a method of WT fault diagnosis based on the Shannon wavelet SVM (SWSVM) and manifold learning. In this method, mixed-domain features are extracted to construct a high-dimensional feature set, manifold learning is used to compress the high-dimensional feature set into low-dimensional eigenvectors, and low-dimensional eigenvectors are inputted into an SWSVM to recognize WT gearbox faults. Gao et al. (2018) proposed a novel fault diagnosis method of WT that combines mean decomposition, multi-scale entropy, least squares, and SVM. In this method, the WT raw vibration signal is divided into several groups for preprocessing. Thereafter, the mean decomposition method is applied to group the signals to obtain the product function. Moreover, the feature parameters are obtained using the multi-scale entropy method of processing the main product function to obtain the feature vector. The characteristic parameters were input into the least squares SVM, which was trained. This method can significantly enhance the fault classification ability of a single SVM and classify the fault type precisely. In the case of single kernel parameters and parameter optimization, Zhao et al. (2018) proposed a fault diagnosis method of WT based on random subspace identification and multi-kernel SVM. Compared with the traditional SVM, the multi-kernel SVM can successfully identify the bearing fault of the WT and has higher fault diagnosis accuracy. In the classification problem, there are not only two classification problems, but also multi classification problems. SVM can also show good classification ability in the face of two classification problems. (Liu K. et al., 2020) used multi-SVM machine to diagnose the fault of renewable energy power grid, which effectively improves the accuracy of fault diagnosis. (Xue et al., 2017). proposed a fault intelligent diagnosis method combining optimal composition of symptom parameters (SPOC) and multi-SVM to diagnose the motor fault, and realized the fault detection and identification of multiple motor faults. In recent years, with the wide application of SVM, experts began to optimize and improve SVM, put forward some machine learning algorithms derived from SVM, put them into the field of fault diagnosis, and achieved good results. (Zhang and Zhou, 2014; Tang et al., 2019).introduced margin mean and margin variance on the basis of SVM and proposed a large margin distributed machine (LDM), and this method has better classification performance than SVM. (Tang et al., 2020a) used LDM to detect the fault of WT's pitch system and optimized it with state transition algorithm (STA), which significantly improved the accuracy of fault detection.

SVM uses inner product kernel function to turn the raw data into linear data through mapping the raw data to a high-dimensional space. However, modeling WT big data is difficult, and the selection of kernel parameters also affects the fault diagnosis accuracy. Moreover, guaranteeing the classification of multi-type WT fault problems is difficult.

Decision Tree

Decision tree (DT) is composed of multiple judgment nodes, and a classification or regression model is formed by the tree structure (Safavian and Landgrebe, 1991). The basic idea is simple, and **Figure 7** shows a WT fault diagnosis model based on DT.

Rabah et al. (Benkercha and Moulahoum, 2018) proposed a fault diagnosis method for a grid-connected WT generator system based on the DT algorithm with high prediction performance and high accuracy. Abdallah et al. (2018) adopted the DT algorithm to perform fault diagnosis on WT, continuously sampled extensive data from thousands of WT at a high rate, and trained integrated DT classifier. Compared with other ML algorithms, DT is easy to implement but it has limitations in dealing with missing values. The WT fault diagnosis process, there are few samples of fault type and more samples of fault-free type. For DT that deals with data

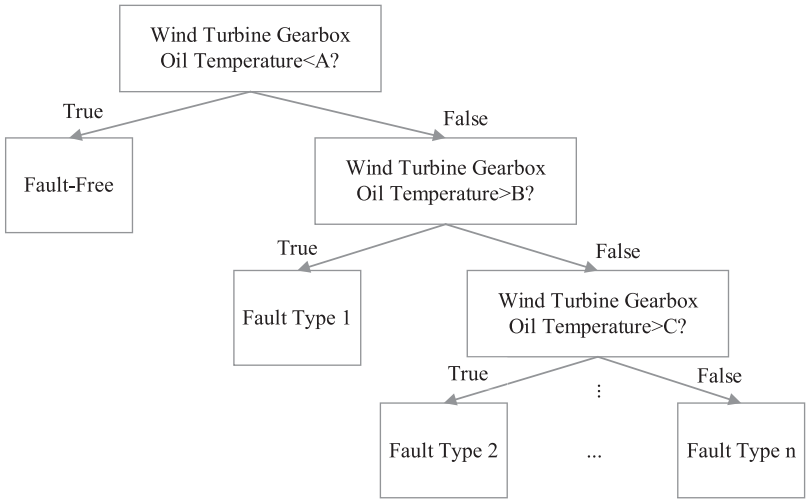

FIGURE 7 | WT fault diagnosis method based on decision tree.

with inconsistent sample sizes in various categories, information gain is biased toward features with additional numerical values, which is easy to overfit and minimally used in WT fault diagnosis.

Ensemble Learning

The basic concept of ensemble learning (Polikar, 2012; Liu et al., 2019) is to adjust and train multiple base learners as ensemble members into a strong learner that should have greater performance on average than any other ensemble member. Thereafter, a model is estabilshed by optimizing the loss function to advance the performance of fault classification. The frequently employed ensemble learning methods include bagging, boosting.

Bootstrap aggregating, also called bagging (Breiman, 1996) applied in regression and statistical classification, is an ML ensemble that obtains a new data set by returning the samples, trains a better base learner based on each new data set, and eventually combines the base learners. The algorithm reduces variance and helps to prevent overfitting. Typical bagging algorithm including random forest (RF). A diagnosis method (Cabrera et al., 2015) was presented for detecting the faults of WT gearboxes, which is based on Random Forest. First, the condition parameters of the vibration signal are extracted by wavelet packet decomposition and used as the input feature of the classification problem. Second, a study approximates the parameter space to find the best mother wave set, and select the best feature through the internal ranking of the random forest classifier. Lastly, the RF algorithm is used to detect the fault of the WT gearbox. To further improve the fault detection rate, Li et al. (2016) proposed a method based on deep RF fusion (DRFF) to improve the fault detection performance of the WT gearboxes. Two deep Boltzmann machines are used to characterize the parameter values of the wavelet packet transform, and the output of the two deep Boltzmann machines is fused into an integrated DRFF model using an RF algorithm. The results indicate that DRFF may improve fault diagnosis capabilities for gearboxes compared with conventional RF.

Boosting (Freund and Schapire, 1996) adjusts the algorithm by giving considerable importance to the bad classification that results in significant improvements in performance of classification. The bagging algorithm focuses on reducing bias facilitates prevention of overfitting. Many algorithms are based on boosting methods, such as XGBoost and LightGBM. Zhang et al. (2018) proposed an efficient WT fault detection method based on the RF and XGBoost. RF is used to rank the features by importance, while XGBoost trains the ensemble classifier for each specific fault based on the top-ranking features. The proposed ensemble classifier can protect against overfitting and experiments verifies the robustness of this method. To enhance the fault diagnosis accuracy, Tang et al. (2020b) proposed the adaptive LightGBM method for the WT gearbox fault detection. The correlation of the WT data samples is analyzed using the maximum information coefficient to realize the feature selection of fault detection. Meanwhile, the LightGBM method after Bayesian hyperparameter optimization is used for the fault detection of WT gearbox. Experiments prove that this method has a low false alarm rate and missing detection rates.

Ensemble learning is widely used in fault diagnosis and early warning of WT with high accuracy. However, some algorithms have slow convergence speed, weak learners rely heavily on one another, and over-fitting problems occur. When using the ensemble methods, the number of iterations, number of base learners, and weights are the issues that should be considered.

Deep Learning

Deep learning was proposed by Hinton et al. (LeCun et al., 2015), and the basic idea is a ML process that includes a multi-level deep network structure through a certain training method based on sample data. Deep learning combines low level features to form a considerably abstract high-level representation to discover the distributed feature representations of data. Deep learning

(Schmidhuber, 2015; Goodfellow et al., 2016) is widely used in image processing, data mining, fault diagnosis (Helbing and Ritter, 2018), and other fields. Different deep learning (Jiang et al., 2018) configurations have also been introduced such as deep belief nets (DBNs), deep auto-encoder (DAE) network, and convolutional neural networks (CNNs).

Toward the WT gearbox faults, Qin et al. (2018) proposed a novel fault diagnosis method that combines DBNs and improved logical Sigmoid unit for the WT gearbox. The integrated approach, which uses the optimized Morlet wavelet transform, kurtosis index, and soft-thresholding is used to extract impulse components from original signals to advance the accuracy of dignosis. Compared with the traditional Sigmoid method, the WT gearbox fault diagnosis method based on deep confidence network and improved logical Sigmoid unit has the higher comprehensive performance. To achieve anomaly diagnosis and fault analysis of WT components, Zhao et al. (2018) proposed a deep learning method based on DAE networks using the WT SCADA data, while the Boltzmann machine builds a deep automatic encoder network model. This method can realize the early warning of the faulty component and derive the physical location of the faulty WT component through the residual of the deep autoencoder network model. Since the diverse operating status of WT with a large amount of noise interference, which leads to a decrease in the accuracy of fault diagnosis of WT. To solve this problem, Chang et al. (2020) proposed a fault diagnosis method for WT based on a concurrent convolution neural network (CeCNN). The raw WT data do not require any prior knowledge, and the characteristics are learned adaptively and directly from input with high accuracy and powerful generalization ability. The convolutional layers of different branches select kernels of varying scales at the same level, thereby improving the accuracy of the WT fault diagnosis. Yi and Jiang (2020) proposed a DAE-based discriminative feature learning for WT blade icing fault detection.

Although deep learning has a strong learning ability and high fault diagnosis accuracy, it requires extensive data and computing power with high cost and high hardware requirements, which are current issues should be considered.

Unsupervised Learning Methods for Wind Turbine Fault Diagnosis

The basic idea of unsupervised learning is the process that a machine learns unlabeled data to reveal the hidden structure, explain the key features of data, and divides them into several categories. Representative technique is clustering. Many algorithms typically used in unsupervised learning are based on the clustering method. Unlike supervised learning methods that analyzes class-labeled instances, unsupervised learning (Figueiredo et al., 2002; Zhang T. and Zhou Z. H., 2018) does not need all information, but trains the information of unlabeled samples. The sample set is clustered according to the similarity between the samples to minimize the intraclass variance and maximize the interclass variance, thereby establishing the model. Unsupervised learning methods can classify and predict test data by extracting hidden concepts and relationships in the data set,

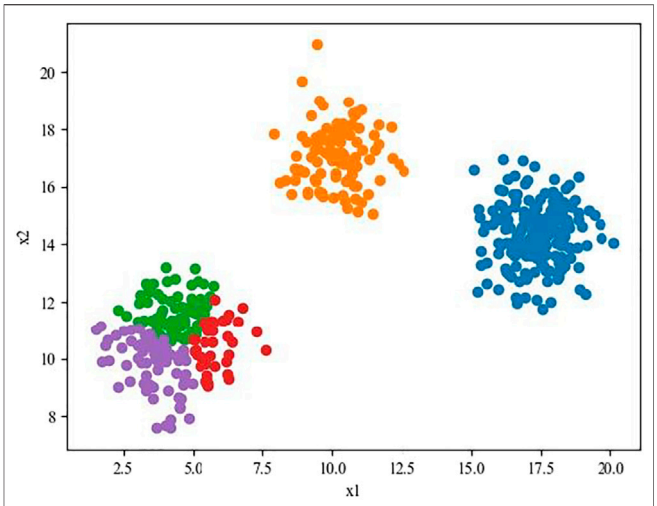

FIGURE 8 | K-means clustering.

which are widely used in fault diagnosis, data mining, and image processing among others. Many methods are typically employed in unsupervised learning, such as the K-means algorithm, fuzzy C-means (FCM), hierarchical clustering method, Gaussian mixture model, and other methods (Hastie et al., 2009).

K-Means

K-means clustering (Kanungo et al., 2002; Jain, 2010) is a simple unsupervised learning method which aims to divide n observations into K clusters, in which the observation belongs to the cluster with the nearest mean. First, we initialize cluster centers and determine K initial points in the data as the center of clustering; Second, we calculate the distance from each point to the center and assign it to the nearest cluster. Third, we recalculate the cluster center to minimize the internal sum. Lastly, the allocation and update operations are repeated until the centers of all clusters no longer change. If all points are allocated to the same cluster as before, then K-means clustering is completed. For example, given the data set of WTs, K-means algorithm was used for clustering and the five types of clustering result is shown in **Figure 8**.

To overcome the sensitivity of K-means to the choice of the initial cluster centers, Yiakopoulos et al. (2011) proposed a K-means clustering method for fault diagnosis of rolling bearings, and the initial centers are selected using features extracted from simulated signals. The fault detection experiments on three types of bearings show that this method can successfully classify faults. Khediri et al. (2012) proposed an unsupervised learning process based on kernel technology, which can separate different non-linear process modes, and effectively detect faults, and reduce the false alarm rate. Kusiak et al. (Kusiak and Verma, 2012) used three different operating curves (i.e., power rotor and blade pitch curves) to monitor the performance of wind farms, and proposed a multivariate outlier based on Mahalanobis distance and K-means clustering. This method, uses the skewness and kurtosis of bivariate data as metrics to evaluate the WT performance, which is simple to apply and has a rapid convergence speed.

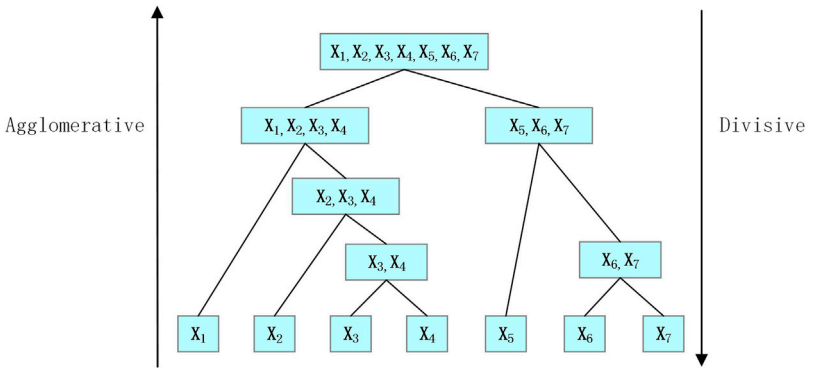

FIGURE 9 | Hierarchical clustering method.

K-means clustering is simple to implement and has a good effect on WT fault diagnosis. However, the choice of the initial cluster center K is difficult to grasp and even cause difficulty in convergence in the case of non-convex data sets. WT has many fault-free samples and few samples of faults, which will result in poor clustering effect when the amount of data is unbalanced.

Fuzzy C-Means

The FCM algorithm (Bezdek et al., 1984; Pal et al., 2005) is a clustering algorithm in which each data point can belong to more than one cluster. The basic idea is to maximize the similarity between objects divided into the same cluster while minimizing the similarity between viaroius clusters.

Given the WT gearboxes fault detection, Luo et al.(Luo and Huang, 2014) proposed a fault diagnosis method based on global local mean decomposition and FCM clustering. In this method, the known sample was clustered using the FCM clustering, and the test sample was classified and recognized, which has simple implementation and good diagnosis results. Although the WT fault diagnosis methods require supervision and training based on historical samples of known faults, collecting samples of known faults is time-consuming and expensive. Given the lack of complete characteristics of known samples in WT, Li et al. (2015) presented a method based on the kernel FCM (KFCM) clustering to the fault diagnosis of the WT gearbox. The KFCM clustering algorithm is used to classify the samples of known samples, and the classification center of each known fault is obtained. Similarity parameters are also calculated to diagnose whether the new data samples belong to the known faults. This method can accurately and effectively diagnose the known and unknown faults of WT.

Some issues should be considered when the FCM algorithm is used in the WT fault diagnosis. For example, a large fault-free sample size and extremely small fault sample size may lead to failure, thereby ensuring that the optimal solution of the fault diagnosis model is found.

Hierarchical Clustering

Hierarchical clustering (Johnson, 1967; Corpet, 1988) is a cluster analysis method in unsupervised learning, which builds a model by establishing a hierarchical structure of clusters. The hierarchical clustering method can be represented as a tree structure (i.e., "tree diagram", which includes roots and leaves. In clustering tree species, the original data points of different categories are the lowest level of the tree, and the top level of the tree is the root node of a cluster. As shown in **Figure 9**, the hierarchical clustering method (Navarro et al., 1997) involves a process that starts from the leaves and successively merges clusters called agglomerative hierarchical clustering; or a process that begins from the root and recursively splits the clusters called divisive hierarchical clustering. The hierarchical clustering method uses Euclidean distance to calculate the distance between the data points of different categories.

Li Y. et al. (2018) proposed a fault diagnosis method based on adaptive multi-scale morphological filters and improved hierarchical arrangement entropy to identify varoius health situations of gearboxes, and used the hierarchical aggregation method to reduce noise fault features extracted from the signal. Liu and Ge (2018) presented a weighted random forest scheme based on hierarchical clustering selection for fault classification in complex industrial processes. The application of the hierarchical clustering method to offline model selection in RF can reduce the complexity of online fault classification.

In the fault diagnosis process of WT, the need to calculate the proximity matrix in the hierarchical clustering algorithm, is time-consuming, and unsuitable for use in the WT big data sets. Hierarchical clustering method is appropriate for the clustering of small data sets, and real-time issues should be considered when dealing with the WT big data.

Gaussian Mixture Model

The Gaussian mixture model (GMM) (Reynolds, 2009) assumes that all data points conform to the Gaussian distribution, and is generated from a mixed finite number of probability models with unknown parameters. GMM can be regarded as the process of fitting a linear combination of multiple Gaussian distribution functions to perform data distribution.

Heyns et al. (2012) proposed a Gaussian hybrid model to detect WT gearbox failures and calculate the negative log-likelihood of the

gearbox bearing vibration signal segment, which represents the healthy gearbox. This method is suitable for nonlinear and non-stationary wind turbine gearbox vibration signals. Given the highly complex and unstable operating conditions of WT, Dong et al. (2013) proposed a multi-parameter WT health assessment framework that considers dynamic operating conditions. After the characteristic parameter selection and GMM based multi-regime modeling, the operation status of WT can be evaluated, which can effectively detect WT faults. In response to frequent WT faults, Luis et al. (Avendaño-Valencia et al., 2017) proposed a fault diagnosis method for WT based on a GMM random coefficient model. The vibration response signals of WT that change with time under the environment and operating conditions are extracted and the model coefficients are determined through the GMM random coefficient framework. The method offering significant performance improvements and most fault levels and types are represented to be correctly diagnosed.

GMM is effective in handling the big volume of WT data samples, but it has a large calculation amount and slow convergence. Selecting the number of sub-models in advance is difficult and is sensitive to abnormal points. When processing small data set samples in WT, the result cannot meet the requirements.

Semi-supervised Learning Methods for Wind Turbine Fault Diagnosis

Semi-supervised learning (Chapelle et al., 2009; Zhou et al., 2014) is a learning paradigm that detects some common features of labeled data samples and unlabeled data samples to help determine the model characteristics and to disseminate labels from labeled data to unlabeled ones, which is an ML method between supervised and unsupervised learning. In selecting data sets, combining unlabeled samples and labeled samples in the training process can improve training accuracy. There are four mainstream paradigms for semi-supervised learning (Zhu, 2005), are the semi-supervised SVM (S3VM), generative model-based, disagreement-based, and graph-based methods. Disagreement-based semi-supervised learning (Blum and Mitchell, 1998) started with the work on co-training (Zhou and Li, 2005) by Blum, which is less affected by the non-convexity of the loss function and the data size and is mainly used in the field of human-computer interaction. The graph-based method (Camps-Valls et al., 2007) was developed by the graph min-cut method (Blum and Chawla, 2001) proposed by Blum, but it is rarely seen in WT fault diagnosis. The method based on S3VM and generative model is also applied in WT fault diagnosis.

S3VM

S3VM (Bennett and Demiriz, 1998) involves the development of SVM in semi-supervised learning (Wang et al., 2017). The major idea of S3VM is to mark unlabeled samples to maximize the interval after the hyperplane is divided. The frequently used S3VM is a transductive SVM (TSVM). The basic idea of this method can be presented as five steps. The first step involves training an SVM classifier with labeled samples. The second step entails using SVM to predict the classification results of unlabeled data. The third step aims to find the opposite label in the predicted unlabeled data that may be wrong for the labeled sample to swap the label, and use the existing labeled sample and unlabeled sample to retrain SVM; The fourth step involves repeating the second and third steps until the best S3VM classifier is obtained. The fifth step entails using use the S3VM classifier to label the unlabeled samples and predict the classification results.

The S3VM methods are widely used in the field of WT fault diagnosis. Liu C. et al. (2020) proposed a fault diagnosis method for rolling bearing based on S3VM using only a few labeled samples to build a model with good classification effect. In order to reduce false alarm rates and improve the discriminative ability of incipient fault features, Mao et al. (2020) proposed an online method for early fault detection of bearings using a semi-supervised architecture. A safe semi-supervised SVM (S4VM) is introduced to identify the sequentially arrived data of the goal bearing as anomalous or normals and fault states and a stacked noise reduction automatic encoder is used to extract depth features from the normal state data and fault state data of the bearing. According to the S4VM generalization error upper bound to adaptively identify the occurrence of an incipient fault. Optimal margin distribution learning machine (ODM), which is also classified based on split hyperplane, has also appeared semi-supervised ODM(ssODM) in recent years, and has been applied to wind turbine fault detection with good performance (Zhang T. and Zhou Z.-H., 2018).

S3VM predicts the unlabeled samples, adds the prediction results to the labeled data set and improves the fault diagnosis rate. However, S3VM should determine a few known WT data samples as a guide. Accordingly, we cannot ensure such WT data samples with delicate information and it is uncertain to know how many WT data samples are needed to achieve an effective S3VM model.

Generative Models

The main idea of the semi-supervised generative model is that the probability that unlabeled samples belong to each category as a set of missing parameters. Thereafter, the expectation maximization (EM) algorithm is used to perform maximum likelihood estimate on the parameters of the generated model. Generative model methods (Zhu, 2005; Kingma et al., 2014) include mixed Gaussian distribution, mixture multinomial distribution, and hidden Markov model.

Ge et al.(Xin et al., 2018) proposed a semi-automatic fault detection method based on a probabilistic model in the form of a hybrid Gaussian with good robustness. Wang et al. (Wang et al., 2015) proposed a comprehensive method based on semi-supervised learning, using a small amount of labeled data and a large amount of unlabeled process data to construct a neighborhood weighted graph. By solving the optimization problem, the optimal regression function and the optimal prediction label matrix of unlabeled data are acquired. This method can obtain the promised results of fault detection and fault diagnosis in the monitoring process. To achieve automatic

detections, Omid et al. (Geramifard et al., 2013) introduced a semi-parametric method based on the hidden Markov model for fault detection and diagnosis of synchronous motors. After training the hidden Markov model classifier (parameter stage), which is based on each probabilistic (non-parametric stage) hidden Markov model. Moreover, the probabilistic inference are used to compute two matrices to solve the efficiency problem in the fault classification process. Li X. et al. (2018) presented a fault detection method on a multivariate Bayesian control scheme and a hidden semi-Markov model to predict early bearing failures of gearboxes. The method of using the continuous-time hidden semi-Markov decision process to characterize the failure process of the gearbox bearing system, which can predict the early failure of the gearbox bearing and detect the remaining useful life at each sampling epoch.

The semi-supervised generative model method has good robustness, but the fault diagnosis model has low accuracy, long model training time, and many iterations. These issues must be considered in WT fault diagnosis.

CONCLUSION AND PERSPECTIVES

Given the rapid development of early wind power generation, wind power equipment has entered a high failure period, and the fault diagnosis methods of WT have high requirements for their operation and maintenance stability. Accordingly, the development fault prediction, fault diagnosis, fault detection, and condition monitoring of WT have improved. Various studies have proposed various methods and strategies for the fault diagnosis and detection of various WT components (Faiz and Moosavi, 2016). Following studies and research on the most recent WT fault diagnosis methods, the current study gathers a review of WT fault diagnosis methods and techniques based onML. Given the many uncertainties in the WT operation, many issues should continue to be considered in the ML-based fault diagnosis of WT.

Improvement of ML algorithms effects. Many types of ML algorithms have advantages and disadvantages, in which among the research fields in the future include improving algorithm performance, optimizing algorithm parameters, combining algorithms, and studying new algorithms. Given that the algorithm has advantages and disadvantages, the need to adopt the advantages and bypass the disadvantages in the algorithm have become urgent issues to be addressed. Moreover, a single algorithm cannot detect all WT faults. Hence, the combined algorithm will become a hot research topic in the future. The advantages and disadvantages of existing algorithms indicated that future research involves proposing and improving new algorithms.

Comprehensive simulation of WT fault conditions. The wind power generation system is a typical complex system, given the uncertain severity and probability of faults. In the research on the WT fault mechanism, only a single fault is modeled, and the single component faults of WT are accompanied by multiple faults, which causes serious damage to WT. All WT units are interconnected and their variables are highly coupled. The occurrence of a fault in a particular component affects all remaining units. Therefore, additional compound fault models should be established to conduct a comprehensive analysis of the WT system.

Research on the feature selection method. WT have many characteristic parameters because the operation state of WT is time-varying Redundant and useless feature parameters will inevitably exist in WT feature extraction. Given the need to extract additional fault features, the research on optimized feature extraction algorithms will become popular in the future, thereby enabling us to better describe and detect the status of WT.

Multi-parameter information fusion. A single sensor or piece of single parameter information cannot acquire dedicated WT operating status information, which entails difficulty in accurately reflecting the fault or normal state of each WT component. Therefore, a multi-parameter information fusion method is adopted to obtain additional parameter information from multiple sensors and improve the efficiency of fault diagnosis.

Establishment of remote WT fault diagnosis system. The WT fault diagnosis system should be able to predict fault and provide the period plan maintenance to keep the WT minimum downtime and maintain long-distance condition monitoring. Long-term available historical data should be provided by the WT fault diagnosis system to set the correct alarm for preventive maintenance. In large wind farms, multiple wind power generation systems must be installed for fault diagnosis and early warning. The need to develop a low-cost and high-efficient remote WT fault diagnosis system should also be considered in the future.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication.

FUNDING

This work was supported in part by the National Natural Science Foundation of China (Grant Nos. 62173050 and 61403046), the Natural Science Foundation of Hunan Province, China (Grant No. 2019JJ40304), Changsha University of Science and Technology "The Double First Class University Plan" International Cooperation and Development Project in Scientific Research in 2018 (Grant No. 2018IC14), the Research Foundation of the Education Bureau of Hunan Province (Grant No.19K007), Hunan Provincial Department of Transportation 2018 Science and Technology Progress and Innovation Plan Project (Grant No. 201843), Energy Conservation and Emission Reduction Hunan University Student Innovation and Entrepreneurship Education Center, Innovative Team of Key Technologies of Energy Conservation, Emission Reduction and Intelligent Control for Power-Generating Equipment and System, CSUST, Hubei Superior

and Distinctive Discipline Group of Mechatronics and Automobiles (XKQ2021003 and XKQ2021010), Major Fund Project of Technical Innovation in Hubei (Grant No. 2017AAA133), and Guangxi Key Laboratory of Trusted Software (No.201728), Graduate Scientific Research Innovation Project of Changsha University of Science and Technology (No. 2021-89).

REFERENCES

Abdallah, I., Dertimanis, V., Mylonas, H., Tatsis, K., Chatzi, E., Dervili, N., et al. (2018). "Fault Diagnosis of Wind Turbine Structures Using Decision Tree Learning Algorithms with Big Data," in 2018 European Safety and Reliability Conference (ESREL 2018), Trondheim, Norway, 3053–3061. doi:10.1201/9781351174664-382

Agatonovic-Kustrin, S., Beresford, R., and analysis, b. (2000). Basic Concepts of Artificial Neural Network (ANN) Modeling and its Application in Pharmaceutical Research. *J. Pharm. Biomed. Anal.* 22 (5), 717–727. doi:10.1016/S0731-7085(99)00272-1

Avendaño-Valencia, L. D., Fassois, S. D., and Processing, S. (2017). Damage/fault Diagnosis in an Operating Wind Turbine under Uncertainty via a Vibration Response Gaussian Mixture Random Coefficient Model Based Framework. *Mech. Syst. Signal Process.* 91, 326–353. doi:10.1016/j.ymssp.2016.11.028

Benkercha, R., and Moulahoum, S. (2018). Fault Detection and Diagnosis Based on C4.5 Decision Tree Algorithm for Grid Connected PV System. *Solar Energy* 173, 610–634. doi:10.1016/j.solener.2018.07.089

Bennett, K. P., and Demiriz, A. (1998). "Semi-supervised Support Vector Machines," in *Advances in Neural Information Processing Systems*, 368–374.)

Bezdek, J. C., Ehrlich, R., and Full, W. J. (1984). FCM: The Fuzzy C-Means Clustering Algorithm. *Comput. Geosciences* 10 (2-3), 191–203. doi:10.1016/0098-3004(84)90020-7

Bielecki, A., Barszcz, T., Wójcik, M., and Bielecka, M. (2014). "Hybrid System of ART and RBF Neural Networks for Classification of Vibration Signals and Operational States of Wind Turbines," in International Conference on Artificial Intelligence and Soft Computing (Springer), 3–11. doi:10.1007/978-3-319-07173-2_1

Blum, A., and Chawla, S. (2001). "Learning from Labeled and Unlabeled Data Using Graph Mincuts," in ICML '01: Proceedings of the Eighteenth International Conference on Machine Learning, 19–26.

Blum, A., and Mitchell, T. (1998). "Combining Labeled and Unlabeled Data with Co-training," in Proceedings of the eleventh annual conference on Computational learning theory, 92–100. doi:10.1145/279943.279962

Breiman, L. (1996). Bagging Predictors. *Mach Learn.* 24 (2), 123–140. doi:10.1007/BF00058655

Cabrera, D., Sancho, F., Sánchez, R.-V., Zurita, G., Cerrada, M., Li, C., et al. (2015). Fault Diagnosis of spur Gearbox Based on Random forest and Wavelet Packet Decomposition. *Front. Mech. Eng.* 10 (3), 277–286. doi:10.1007/s11465-015-0348-8

Camps-Valls, G., Bandos Marsheva, T. V., and Zhou, D. (2007). Semi-supervised Graph-Based Hyperspectral Image Classification. *IEEE Trans. Geosci. Remote Sensing* 45 (10), 3044–3054. doi:10.1109/TGRS.2007.895416

Chang, Y., Chen, J., Qu, C., and Pan, T. (2020). Intelligent Fault Diagnosis of Wind Turbines via a Deep Learning Network Using Parallel Convolution Layers with Multi-Scale Kernels. *Renew. Energ.* 153, 205–213. doi:10.1016/j.renene.2020.02.004

Chapelle, O., Scholkopf, B., and Zien, Eds., A. (2009). Semi-Supervised Learning. *IEEE Trans. Neural Netw.* 20 (3), 542. doi:10.1109/TNN.2009.2015974

Chen, J., Li, Z., Pan, J., Chen, G., Zi, Y., Yuan, J., et al. (2016a). Wavelet Transform Based on Inner Product in Fault Diagnosis of Rotating Machinery: A Review. *Mech. Syst. Signal Process.* 70-71, 1–35. doi:10.1016/j.ymssp.2015.08.023

Chen, J., Pan, J., Li, Z., Zi, Y., and Chen, X. (2016b). Generator Bearing Fault Diagnosis for Wind Turbine via Empirical Wavelet Transform Using Measured Vibration Signals. *Renew. Energ.* 89, 80–92. doi:10.1016/j.renene.2015.12.010

Cho, S., Gao, Z., and Moan, T. (2018). Model-based Fault Detection, Fault Isolation and Fault-Tolerant Control of a Blade Pitch System in Floating Wind Turbines. *Renew. Energ.* 120, 306–321. doi:10.1016/j.renene.2017.12.102

Clifton, A., Kilcher, L., Lundquist, J. K., and Fleming, P. (2013). Using Machine Learning to Predict Wind Turbine Power Output. *Environ. Res. Lett.* 8 (2), 024009. doi:10.1088/1748-9326/8/2/024009

Corpet, F. (1988). Multiple Sequence Alignment with Hierarchical Clustering. *Nucleic Acids Res.* 16 (22), 10881–10890. doi:10.1093/nar/16.22.10881

Dao, P. B., Staszewski, W. J., Barszcz, T., and Uhl, T. (2018). Condition Monitoring and Fault Detection in Wind Turbines Based on Cointegration Analysis of SCADA Data. *Renew. Energ.* 116, 107–122. doi:10.1016/j.renene.2017.06.089

de Azevedo, H. D. M., Araújo, A. M., and Bouchonneau, N. (2016). A Review of Wind Turbine Bearing Condition Monitoring: State of the Art and Challenges. *Renew. Sustainable Energ. Rev.* 56, 368–379. doi:10.1016/j.rser.2015.11.032

Ding, S. X. (2012). Data-driven Design of Model-Based Fault Diagnosis Systems. *IFAC Proc. Volumes* 45 (15), 840–847. doi:10.3182/20120710-4-SG-2026.00105

Dong, Y., Fang, F., and Gu, Y. (2013). Dynamic Evaluation of Wind Turbine Health Condition Based on Gaussian Mixture Model and Evidential Reasoning. *J. Renew. Sustainable Energ.* 5 (3), 033117. doi:10.1063/1.4808018

Du, Z., Chen, X., Zhang, H., and Yan, R. (2015). Sparse Feature Identification Based on union of Redundant Dictionary for Wind Turbine Gearbox Fault Diagnosis. *IEEE Trans. Ind. Electron.* 62 (10), 6594–6605. doi:10.1109/TIE.2015.2464297

Elizondo, J. L., Rivera, M., and Wheeler, P. (2019). "Wind Energy Development and Technology in the World: A Brief Overview," in 2019 IEEE CHILEAN Conference on Electrical, Electronics Engineering, Information and Communication Technologies (CHILECON) (IEEE), 1–5. doi:10.1109/CHILECON47746.2019.8987714

Faiz, J., and Moosavi, S. M. M. (2016). Eccentricity Fault Detection - from Induction Machines to DFIG-A Review. *Renew. Sustainable Energ. Rev.* 55, 169–179. doi:10.1016/j.rser.2015.10.113

Figueiredo, M. A. T., Jain, A. K., and intelligence, m. (2002). Unsupervised Learning of Finite Mixture Models. *IEEE Trans. Pattern Anal. Mach. Intell.* 24 (3), 381–396. doi:10.1109/34.990138

Freund, Y., and Schapire, R. E. (1996). "Experiments with a New Boosting Algorithm," in Proceedings of the Thirteenth International Conference on International Conference on Machine Learning, 148–156.

Gao, Q. W., Liu, W. Y., Tang, B. P., and Li, G. J. (2018). A Novel Wind Turbine Fault Diagnosis Method Based on Intergral Extension Load Mean Decomposition Multiscale Entropy and Least Squares Support Vector Machine. *Renew. Energ.* 116, 169–175. doi:10.1016/j.renene.2017.09.061

Gao, Z., Cecati, C., and Ding, S. X. (2015). A Survey of Fault Diagnosis and Fault-Tolerant Techniques-Part I: Fault Diagnosis with Model-Based and Signal-Based Approaches. *IEEE Trans. Ind. Electron.* 62 (6), 3757–3767. doi:10.1109/TIE.2015.2417501

Garg, H., and Dahiya, R. (2017). Current Signature Analysis and its Application in the Condition Monitoring of Wind Turbine for Rotor Faults. *Energy Syst.* 8 (3), 495–510. doi:10.1007/s12667-016-0208-6

Geramifard, O., Xu, J.-X., and Kumar Panda, S. (2013). Fault Detection and Diagnosis in Synchronous Motors Using Hidden Markov Model-Based Semi-nonparametric Approach. *Eng. Appl. Artif. Intelligence* 26 (8), 1919–1929. doi:10.1016/j.engappai.2013.06.005

Goodfellow, I., Bengio, Y., and Courville, A. (2016). *Deep Learning*. MIT press.

Habibi, H., Howard, I., and Simani, S. (2019). Reliability Improvement of Wind Turbine Power Generation Using Model-Based Fault Detection and Fault Tolerant Control: A Review. *Renew. Energ.* 135, 877–896. doi:10.1016/j.renene.2018.12.066

Habibi, H., Rahimi Nohooji, H., and Howard, I. (2017). Power Maximization of Variable-Speed Variable-Pitch Wind Turbines Using Passive Adaptive Neural Fault Tolerant Control. *Front. Mech. Eng.* 12 (3), 377–388. doi:10.1007/s11465-017-0431-4

Hahn, B., Durstewitz, M., and Rohrig, K. (2007). "Reliability of Wind Turbines," in *Wind Energy* (Springer), 329–332. doi:10.1007/978-3-540-33866-6_62

Hastie, T., Tibshirani, R., and Friedman, J. (2009). "Unsupervised Learning," in *The Elements of Statistical Learning* (Springer), 485–585. doi:10.1007/978-0-387-84858-7_14

Helbing, G., and Ritter, M. (2018). Deep Learning for Fault Detection in Wind Turbines. *Renew. Sustainable Energ. Rev.* 98, 189–198. doi:10.1016/j.rser.2018.09.012

Heyns, T., Heyns, P. S., and De Villiers, J. P. (2012). Combining Synchronous Averaging with a Gaussian Mixture Model novelty Detection Scheme for Vibration-Based Condition Monitoring of a Gearbox. *Mech. Syst. Signal Process.* 32, 200–215. doi:10.1016/j.ymssp.2012.05.008

Himani, R., and Dahiya, R. (2018). Condition Monitoring of Wind Turbine for Rotor Fault Detection under Non Stationary Conditions. *Ain Shams Eng. J.* 9 (4), 2441–2452. doi:10.1016/j.asej.2017.04.002

Hongshan, Z., Gao, Y., Liu, H., Li, L., and Energy, C. (2019). Fault Diagnosis of Wind Turbine Bearing Based on Stochastic Subspace Identification and Multi-Kernel Support Vector Machine. *J. Mod. Power Syst. Clean. Energ.* 7 (2), 350–356. doi:10.1007/s40565-018-0402-8

Hossain, M. M., Ali, M. H., and reviews, S. e. (2015). Future Research Directions for the Wind Turbine Generator System. *Renew. Sustainable Energ. Rev.* 49, 481–489. doi:10.1016/j.rser.2015.04.126

Igba, J., Alemzadeh, K., Durugbo, C., and Eiriksson, E. T. (2016). Analysing RMS and Peak Values of Vibration Signals for Condition Monitoring of Wind Turbine Gearboxes. *Renew. Energ.* 91, 90–106. doi:10.1016/j.renene.2016.01.006

Jain, A. K. (2010). Data Clustering: 50 Years beyond K-Means. *Pattern recognition Lett.* 31 (8), 651–666. doi:10.1016/j.patrec.2009.09.011

Jiang, G., He, H., Yan, J., and Xie, P. (2019). Multiscale Convolutional Neural Networks for Fault Diagnosis of Wind Turbine Gearbox. *IEEE Trans. Ind. Electron.* 66 (4), 3196–3207. doi:10.1109/TIE.2018.2844805

Jiménez, A. A., Muñoz, C. Q. G., Márquez, F. P. G. J. R. E., and Safety, S. (2019). Dirt and Mud Detection and Diagnosis on a Wind Turbine Blade Employing Guided Waves and Supervised Learning Classifiers. *Reliability Engineering% System Saf.* 184, 2–12. doi:10.1016/j.ress.2018.02.013

Johnson, S. C. (1967). Hierarchical Clustering Schemes. *Psychometrika* 32 (3), 241–254. doi:10.1007/BF02289588

Kanungo, T., Mount, D. M., Netanyahu, N. S., Piatko, C. D., Silverman, R., Wu, A. Y., et al. (2002). An Efficient K-Means Clustering Algorithm: Analysis and Implementation. *IEEE Trans. Pattern Anal. Machine Intell.* 24 (7), 881–892. doi:10.1109/TPAMI.2002.1017616

Khediri, I. B., Weihs, C., and Limam, M. (2012). Kernel K-Means Clustering Based Local Support Vector Domain Description Fault Detection of Multimodal Processes. *Expert Syst. Appl.* 39 (2), 2166–2171. doi:10.1016/j.eswa.2011.07.045

Kingma, D. P., Mohamed, S., Rezende, D. J., and Welling, M. (2014). "Semi-supervised Learning with Deep Generative Models," in Advances in neural information processing systems, 3581–3589.

Kusiak, A., and Verma, A. (2013). Monitoring Wind Farms with Performance Curves. *IEEE Trans. Sustain. Energ.* 4 (1), 192–199. doi:10.1109/TSTE.2012.2212470

Lan, J., Patton, R. J., and Zhu, X. (2018). Fault-tolerant Wind Turbine Pitch Control Using Adaptive Sliding Mode Estimation. *Renew. Energ.* 116, 219–231. doi:10.1016/j.renene.2016.12.005

Laouti, N., Sheibat-Othman, N., and Othman, S. (2011). Support Vector Machines for Fault Detection in Wind Turbines. *IFAC Proc. Volumes* 44 (1), 7067–7072. doi:10.3182/20110828-6-IT-1002.02560

Leahy, K., Hu, R. L., Konstantakopoulos, I. C., Spanos, C. J., and Agogino, A. M. (2016). "Diagnosing Wind Turbine Faults Using Machine Learning Techniques Applied to Operational Data," in 2016 IEEE International Conference on Prognostics and Health Management (ICPHM) (IEEE), 1–8. doi:10.1109/ICPHM.2016.7542860

LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep Learning. *nature* 521 (7553), 436–444. doi:10.1038/nature14539

Lei, Y., Yang, B., Jiang, X., Jia, F., Li, N., and Nandi, A. K. (2020). Applications of Machine Learning to Machine Fault Diagnosis: A Review and Roadmap. *Mech. Syst. Signal Process.* 138, 106587. doi:10.1016/j.ymssp.2019.106587

Li, C., Sanchez, R.-V., Zurita, G., Cerrada, M., Cabrera, D., Vásquez, R. E., et al. (2016). Gearbox Fault Diagnosis Based on Deep Random forest Fusion of Acoustic and Vibratory Signals. *Mech. Syst. Signal Process.* 76-77, 283–293. doi:10.1016/j.ymssp.2016.02.007

Li, X., Makis, V., Zuo, H., Cai, J., and Engineering, I. (2018a). Optimal Bayesian Control Policy for Gear Shaft Fault Detection Using Hidden Semi-markov Model. *Comput. Ind. Eng.* 119, 21–35. doi:10.1016/j.cie.2018.03.026

Li, Y., Li, G., Yang, Y., Liang, X., Xu, M., and Processing, S. (2018b). A Fault Diagnosis Scheme for Planetary Gearboxes Using Adaptive Multi-Scale Morphology Filter and Modified Hierarchical Permutation Entropy. *Mech. Syst. Signal Process.* 105, 319–337. doi:10.1016/j.ymssp.2017.12.008

Liang, J., Zhang, K., Al-Durra, A., and Zhou, D. (2020). A Novel Fault Diagnostic Method in Power Converters for Wind Power Generation System. *Appl. Energ.* 266, 114851. doi:10.1016/j.apenergy.2020.114851

Lin, Y., Tu, L., Liu, H., and Li, W. (2016). Fault Analysis of Wind Turbines in China. *Renew. Sustainable Energ. Rev.* 55, 482–490. doi:10.1016/j.rser.2015.10.149

Liu, C., Gryllias, K., and Processing, S. (2020a). A Semi-supervised Support Vector Data Description-Based Fault Detection Method for Rolling Element Bearings Based on Cyclic Spectral Analysis. *Mech. Syst. Signal Process.* 140, 106682. doi:10.1016/j.ymssp.2020.106682

Liu, H., Chen, C., Lv, X., Wu, X., and Liu, M. (2019). Deterministic Wind Energy Forecasting: A Review of Intelligent Predictors and Auxiliary Methods. *Energ. Convers. Management* 195, 328–345. doi:10.1016/j.enconman.2019.05.020

Liu, K., Dong, W., Dong, H., Wei, J., and Xiao, S. (2020b). A Complex Fault Diagnostic Approach of Active Distribution Network Based on SBS-SFS Optimized Multi-SVM. *Math. Probl. Eng.* 2020, 1–12. doi:10.1155/2020/8423571

Liu, W. Y., Tang, B. P., Han, J. G., Lu, X. N., Hu, N. N., He, Z. Z., et al. (2015). The Structure Healthy Condition Monitoring and Fault Diagnosis Methods in Wind Turbines: A Review. *Renew. Sustainable Energ. Rev.* 44, 466–472. doi:10.1016/j.rser.2014.12.005

Liu, X., Gao, Z., and Chen, M. Z. Q. (2017). Takagi-Sugeno Fuzzy Model Based Fault Estimation and Signal Compensation with Application to Wind Turbines. *IEEE Trans. Ind. Electron.* 64 (7), 5678–5689. doi:10.1109/TIE.2017.2677327

Liu, Y., and Ge, Z. (2018). Weighted Random Forests for Fault Classification in Industrial Processes with Hierarchical Clustering Model Selection. *J. Process Control.* 64, 62–70. doi:10.1016/j.jprocont.2018.02.005

Luo, X., and Huang, X. (2014). Fault Diagnosis of Wind Turbine Based on Elmd and Fcm. *Tomej* 8 (1), 716–720. doi:10.2174/1874155X01408010716

Mao, W., Tian, S., Fan, J., Liang, X., and Safian, A. (2020). Online Detection of Bearing Incipient Fault with Semi-supervised Architecture and Deep Feature Representation. *J. Manufacturing Syst.* 55, 179–198. doi:10.1016/j.jmsy.2020.03.005

MUA. (2017). *Global Wind Report*. Brussels: Global Wind Energy Council.

Nasiri, M., Milimonfared, J., Fathi, S. H., and Reviews, S. E. (2015). A Review of Low-Voltage Ride-Through Enhancement Methods for Permanent Magnet Synchronous Generator Based Wind Turbines. *Renew. Sustainable Energ. Rev.* 47, 399–415. doi:10.1016/j.rser.2015.03.079

Navarro, J. F., Frenk, C. S., and White, S. D. M. (1997). A Universal Density Profile from Hierarchical Clustering. *ApJ* 490 (2), 493–508. doi:10.1086/304888

Ouanas, A., Medoued, A., Mordjaoui, M., Lebaroud, A., and Sayad, D. (2018). Fault Diagnosis in Yaw Drive Induction Motor for Wind Turbine. *Wind Eng.* 42 (6), 576–595. doi:10.1177/0309524X18780379

Pal, N. R., Pal, K., Keller, J. M., and Bezdek, J. C. (2005). A Possibilistic Fuzzy C-Means Clustering Algorithm. *IEEE Trans. Fuzzy Syst.* 13 (4), 517–530. doi:10.1109/TFUZZ.2004.840099

Pei, Y., Qian, Z., Jing, B., Kang, D., and Zhang, L. (2018). Data-driven Method for Wind Turbine Yaw Angle Sensor Zero-point Shifting Fault Detection. *Energies* 11 (3), 553. doi:10.3390/en11030553

Polikar, R. (2012). "Ensemble Learning," in *Ensemble Machine Learning* (Springer), 1–34. doi:10.1007/978-1-4419-9326-7_1

Qiao, W., and Lu, D. (2015). A Survey on Wind Turbine Condition Monitoring and Fault Diagnosis-Part I: Components and Subsystems. *IEEE Trans. Ind. Electron.* 62 (10), 6536–6545. doi:10.1109/TIE.2015.2422112

Qin, S. J. (2012). Survey on Data-Driven Industrial Process Monitoring and Diagnosis. *Annu. Rev. Control* 36 (2), 220–234. doi:10.1016/j.arcontrol.2012.09.004

Qin, Y., Wang, X., and Zou, J. (2019). The Optimized Deep Belief Networks with Improved Logistic Sigmoid Units and Their Application in Fault Diagnosis for Planetary Gearboxes of Wind Turbines. *IEEE Trans. Ind. Electron.* 66 (5), 3814–3824. doi:10.1109/TIE.2018.2856205

Raghavendra. N, S., and Deka, P. C. (2014). Support Vector Machine Applications in the Field of Hydrology: a Review. *Appl. soft Comput.* 19, 372–386. doi:10.1016/j.asoc.2014.02.002

Reynolds, D. (2009). Gaussian Mixture Models. *Encyclopedia of biometrics* 741, 659–663. doi:10.1007/978-0-387-73003-5_196

Safavian, S. R., and Landgrebe, D. (1991). A Survey of Decision Tree Classifier Methodology. *IEEE Trans. Syst. Man. Cybern.* 21 (3), 660–674. doi:10.1109/21.97458

Salameh, J. P., Cauet, S., Etien, E., Sakout, A., Rambault, L., and Processing, S. (2018). Gearbox Condition Monitoring in Wind Turbines: A Review. *Mech. Syst. Signal Process.* 111, 251–264. doi:10.1016/j.ymssp.2018.03.052

Samanta, B., Al-Balushi, K. R., and processing, s. (2003). Artificial Neural Network Based Fault Diagnostics of Rolling Element Bearings Using Time-Domain Features. *Mech. Syst. Signal Process.* 17 (2), 317–328. doi:10.1006/mssp.2001.1462

Saravanan, N., and Ramachandran, K. I. (2010). Incipient Gear Box Fault Diagnosis Using Discrete Wavelet Transform (DWT) for Feature Extraction and Classification Using Artificial Neural Network (ANN). *Expert Syst. Appl.* 37 (6), 4168–4181. doi:10.1016/j.eswa.2009.11.006

Schmidhuber, J. (2015). Deep Learning in Neural Networks: An Overview. *Neural networks* 61, 85–117. doi:10.1016/j.neunet.2014.09.003

Schwenker, F., and Trentin, E. (2014). Pattern Classification and Clustering: A Review of Partially Supervised Learning Approaches. *Pattern Recognition Lett.* 37, 4–14. doi:10.1016/j.patrec.2013.10.017

Silva, J. C. d., Saxena, A., Balaban, E., and Goebel, K. (2012). A Knowledge-Based System Approach for Sensor Fault Modeling, Detection and Mitigation. *Expert Syst. Appl.* 39 (12), 10977–10989. doi:10.1016/j.eswa.2012.03.026

Stetco, A., Dinmohammadi, F., Zhao, X., Robu, V., Flynn, D., Barnes, M., et al. (2019). Machine Learning Methods for Wind Turbine Condition Monitoring: A Review. *Renew. Energ.* 133, 620–635. doi:10.1016/j.renene.2018.10.047

Tang, B., Song, T., Li, F., and Deng, L. (2014). Fault Diagnosis for a Wind Turbine Transmission System Based on Manifold Learning and Shannon Wavelet Support Vector Machine. *Renew. Energ.* 62, 1–9. doi:10.1016/j.renene.2013.06.025

Tang, M., Ding, S. X., Yang, C., Cheng, F., Shardt, Y. A. W., Long, W., et al. (2019). Cost-sensitive Large Margin Distribution Machine for Fault Detection of Wind Turbines. *Cluster Comput.* 22, 7525–7537. doi:10.1007/s10586-018-1854-3

Tang, M., Hu, J., Kuang, Z., Wu, H., Zhao, Q., and Peng, S. (2020a). Fault Detection of the Wind Turbine Variable Pitch System Based on Large Margin Distribution Machine Optimized by the State Transition Algorithm. *Math. Probl. Eng.* 2020, 1–9. doi:10.1155/2020/9718345

Tang, M., Zhao, Q., Ding, S. X., Wu, H., Li, L., Long, W., et al. (2020b). An Improved LightGBM Algorithm for Online Fault Detection of Wind Turbine Gearboxes. *Energies* 13 (4), 807. doi:10.3390/en13040807

Teng, W., Ding, X., Zhang, X., Liu, Y., and Ma, Z. (2016). Multi-fault Detection and Failure Analysis of Wind Turbine Gearbox Using Complex Wavelet Transform. *Renew. Energ.* 93, 591–598. doi:10.1016/j.renene.2016.03.025

Toubakh, H., and Sayed-Mouchaweh, M. (2016). Hybrid Dynamic Classifier for Drift-like Fault Diagnosis in a Class of Hybrid Dynamic Systems: Application to Wind Turbine Converters. *Neurocomputing* 171, 1496–1516. doi:10.1016/j.neucom.2015.07.073

Vapnik, V. (2013). *The Nature of Statistical Learning Theory*. Springer science & business media.

Wang, J., Peng, Y., and Qiao, W. (2016a). Current-aided Order Tracking of Vibration Signals for Bearing Fault Diagnosis of Direct-Drive Wind Turbines. *IEEE Trans. Ind. Electron.* 63 (10), 6336–6346. doi:10.1109/TIE.2016.2571258

Wang, L., Zhang, Z., Long, H., Xu, J., and Liu, R. (2017b). Wind Turbine Gearbox Failure Identification with Deep Neural Networks. *IEEE Trans. Ind. Inf.* 13 (3), 1360–1368. doi:10.1109/TII.2016.2607179

Wang, T., Han, Q., Chu, F., and Feng, Z. (2019). Vibration Based Condition Monitoring and Fault Diagnosis of Wind Turbine Planetary Gearbox: A Review. *Mech. Syst. Signal Process.* 126, 662–685. doi:10.1016/j.ymssp.2019.02.051

Wang, W., Xi, J., Chong, A., and Li, L. (2017). Driving Style Classification Using a Semisupervised Support Vector Machine. *IEEE Trans. Human-mach. Syst.* 47 (5), 650–660. doi:10.1109/THMS.2017.2736948

Wang, X., Feng, H., and Fan, Y. (2015). Fault Detection and Classification for Complex Processes Using Semi-supervised Learning Algorithm. *Chemometrics Intell. Lab. Syst.* 149, 24–32. doi:10.1016/j.chemolab.2015.10.019

Xi, L., Wu, J., Xu, Y., Sun, H., and Systems, L. (2021). Automatic Generation Control Based on Multiple Neural Networks with Actor-Critic Strategy. *IEEE Trans. Neural Netw. Learn. Syst.* 32 (6), 2483–2493. doi:10.1109/TNNLS.2020.3006080

Xin, G., Hamzaoui, N., and Antoni, J. (2018). Semi-automated Diagnosis of Bearing Faults Based on a Hidden Markov Model of the Vibration Signals. *Measurement* 127, 141–166. doi:10.1016/j.measurement.2018.05.040

Xue, H., Wang, M., Li, Z., and Chen, P. (2017). Sequential Fault Detection for Sealed Deep Groove ball Bearings of In-Wheel Motor in Variable Operating Conditions. *J. Vibroeng* 19 (8), 5947–5959. doi:10.21595/jve.2017.18413

Yang, B., Liu, R., and Chen, X. (2017). Fault Diagnosis for a Wind Turbine Generator Bearing via Sparse Representation and Shift-Invariant K-SVD. *IEEE Trans. Ind. Inf.* 13 (3), 1321–1331. doi:10.1109/TII.2017.2662215

Yang, T., Pen, H., Wang, Z., and Chang, C. S. (2016). Feature Knowledge Based Fault Detection of Induction Motors through the Analysis of Stator Current Data. *IEEE Trans. Instrum. Meas.* 65 (3), 549–558. doi:10.1109/TIM.2015.2498978

Yang, X., Li, J., Liu, W., and Guo, P. (2011). Petri Net Model and Reliability Evaluation for Wind Turbine Hydraulic Variable Pitch Systems. *Energies* 4 (6), 978–997. doi:10.3390/en4060978

Yi, H., and Jiang, Q. (2020). Discriminative Feature Learning for Blade Icing Fault Detection of Wind Turbine. *Meas. Sci. Technol.* 31, 115102. doi:10.1088/1361-6501/ab9bb8

Yiakopoulos, C. T., Gryllias, K. C., and Antoniadis, I. A. (2011). Rolling Element Bearing Fault Detection in Industrial Environments Based on a K-Means Clustering Approach. *Expert Syst. Appl.* 38 (3), 2888–2911. doi:10.1016/j.eswa.2010.08.083

Yin, S., Wang, G., and Karimi, H. R. (2014). Data-driven Design of Robust Fault Detection System for Wind Turbines. *Mechatronics* 24 (4), 298–306. doi:10.1016/j.mechatronics.2013.11.009

Zappalá, D., Sarma, N., Djurović, S., Crabtree, C. J., Mohammad, A., and Tavner, P. J. (2019). Electrical & Mechanical Diagnostic Indicators of Wind Turbine Induction Generator Rotor Faults. *Renew. Energ.* 131, 14–24. doi:10.1016/j.renene.2018.06.098

Zeng, R., Xu, L., Yao, L., Finney, S. J., and Wang, Y. (2016). Hybrid HVDC for Integrating Wind Farms with Special Consideration on Commutation Failure. *IEEE Trans. Power Deliv.* 31 (2), 789–797. doi:10.1109/TPWRD.2015.2440354

Zhang, D., Qian, L., Mao, B., Huang, C., Huang, B., and Si, Y. (2018). A Data-Driven Design for Fault Detection of Wind Turbines Using Random Forests and XGboost. *IEEE Access* 6, 21020–21031. doi:10.1109/ACCESS.2018.2818678

Zhang, T., and Zhou, Z.-H. (2014). "Large Margin Distribution Machine," in Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining 2014 (New York, New York, USA: Association for Computing Machinery), 313–322. doi:10.1145/2623330.2623710

Zhang, T., and Zhou, Z.-H. (2018b). "Semi-Supervised Optimal Margin Distribution Machines," in Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence: IJCAI 2018, 3104–3110. doi:10.24963/ijcai.2018/431

Zhang, T., and Zhou, Z. H. (2018a). "Optimal Margin Distribution Clustering," in Proceedings of the AAAI Conference on Artificial Intelligence.

Zhang, Y., Lu, W., and Chu, F. (2017). Planet Gear Fault Localization for Wind Turbine Gearbox Using Acoustic Emission Signals. *Renew. Energ.* 109, 449–460. doi:10.1016/j.renene.2017.03.035

Zhang, Z.-Y., and Wang, K.-S. (2014). Wind Turbine Fault Detection Based on SCADA Data Analysis Using ANN. *Adv. Manuf.* 2 (1), 70–78. doi:10.1007/s40436-014-0061-6

Zhao, H., Liu, H., Hu, W., and Yan, X. (2018). Anomaly Detection and Fault Analysis of Wind Turbine Components Based on Deep Learning Network. *Renew. Energ.* 127, 825–834. doi:10.1016/j.renene.2018.05.024

Zhao, L., Zuowei Pan, Z., Changsheng Shao, C., and Qianzhi Yang, Q. (2015). Application of SOM Neural Network in Fault Diagnosis of Wind Turbine. *IET*. doi:10.1049/cp.2015.0446

Zhong, M., Xue, T., and Ding, S. X. (2018). A Survey on Model-Based Fault Diagnosis for Linear Discrete Time-Varying Systems. *Neurocomputing* 306, 51–60. doi:10.1016/j.neucom.2018.04.037

Zhou, S., Chen, Q., and Wang, X. (2014). Fuzzy Deep Belief Networks for Semi-supervised Sentiment Classification. *Neurocomputing* 131, 312–322. doi:10.1016/j.neucom.2013.10.011

Zhou, Z.-H. (2018). A Brief Introduction to Weakly Supervised Learning. *Natl. Sci. Rev.* 5 (1), 44–53. doi:10.1093/nsr/nwx106

Zhou, Z. H., and Li, M. (2005). "Semi-Supervised Regression with Co-training," in Proceedings of the 19th international joint conference on Artificial intelligence IJCAI, 908–913.

Zhu, X. J. (2005). *Semi-supervised Learning Literature Survey*. University of Wisconsin-Madison Department of Computer Sciences.

Zhuang Li, Z., Rui Jiang, R., Zhiyong Ma, Z., and Yibing Liu, Y. (2015). "Fault Diagnosis of Wind Turbine Gearbox Based on Kernel Fuzzy C-Means Clustering," in International Conference on Renewable Power Generation, 1–4. doi:10.1049/cp.2015.0418

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Energy Research

ORIGINAL RESEARCH
published: 07 December 2021
doi: 10.3389/fenrg.2021.709708

Adaptive IES Load Forecasting Method Based on the Octopus Model

Na Zhang, Xiao Pan, Yihe Wang, Mingli Zhang, Mengzeng Cheng and Wenying Shang*

Economic and Technological Research Institute of State Grid Liaoning Electric Power Co., Ltd, Shenyang, China

Improving the accuracy and speed of integrated energy system load forecasting is a great significance for improving the real-time scheduling and optimized operation of the integrated energy system. In order to achieve rapid and accurate forecasting of the integrated energy system, this paper proposes an adaptive integrate energy system (IES) load forecasting method based on the octopus model. This method uses long short-term memory (LSTM), support vector machines (SVMs), restricted Boltzmann machines (RBMs), and Elman neural network as the octopus model quadrupeds. Through taking over differences in different data and training principles and utilizing the advantages of the octopus quadruped model, a special octopus-head and XGBoost algorithm were adopted to set the weight of the octopus' quadruped and prevent local minimum points in the model. We train the octopus model through RMSProp adaptive learning algorithm, constrain the learning rate, get the best parameters, and improve the model's adaptability to different types of data. In addition, for the incomplete comprehensive energy load data, the generative confrontation network is used to fill it. The simulation results show that compared with other prediction methods, the effectiveness and feasibility of the method proposed in this paper are verified.

Keywords: comprehensive energy, generative confrontation network, XGBoost algorithm, RMSProp adaptive learning, octopus model

OPEN ACCESS

Edited by:
Yonghao Gui,
Aalborg University, Denmark

Reviewed by:
Chenghao Sun,
Northeastern University, China
Xuguang Hu,
Northeastern University, China

****Correspondence:***
Na Zhang
q2629249062@126.com

Specialty section:
This article was submitted to Smart Grids, a section of the journal Frontiers in Energy Research

Received: *14 May 2021*
Accepted: *28 June 2021*
Published: *07 December 2021*

Citation:
Zhang N, Pan X, Wang Y, Zhang M, Cheng M and Shang W (2021) Adaptive IES Load Forecasting Method Based on the Octopus Model. Front. Energy Res. 9:709708. doi: 10.3389/fenrg.2021.709708

INTRODUCTION

An integrate energy system (IES) (AlDahoul et al., 2021) refers to the reasonable distribution and utilization of electricity, gas, heat, wind, and other energies within a certain range or area, using professional technology and operating modes, to achieve energy efficiency, interaction, and complementarity. While ensuring that the energy consumption of individual users meets their needs, it also improves the energy consumption of the entire range and improves the overall energy utilization efficiency.

At present, the scale of our country's integrated energy system continues to expand, and the load forecast of the integrated energy system affects the operation and planning of the integrated energy system. Improving the speed and accuracy of load forecasting of the integrated energy system is of vital importance for realizing the real-time scheduling and optimized operation of the integrated energy system. In recent years, scholars at home and abroad have conducted some research on load forecasting; Bian et al. (2020) built gray correlation analysis under the LSTM (Dai and Zhao, 2020) neural network model, improved the traditional LSTM neural network's processing method of time series and nonlinear data, and improved the accuracy of short-term load forecasting, but did not consider the comprehensive energy relationship. Daniel et al. (2020) used the packet decomposition and cyclic neural network to decompose the electric, cold, and heat loads in frequency bands and determined the prediction method by judging the correlation between each frequency band. This

method can accurately predict the loads with strong and weak autocorrelation, but does not consider the impact on incomplete data onto the prediction accuracy. Dong et al. (2021) used Copula theory to analyze the correlation between the comprehensive energy load, established a time series, and used the K-means clustering algorithm to design a radial basis function neural network. The calculation method of the model is simple and easy to design, but the prediction accuracy of some outlier data points and isolated data points is low. Guo et al. (2021) considered that the training process of the traditional wavelet neural network (WNN) (Guo et al., 2020) is prone to the shortcoming of too fast convergence speed and proposed a WNN prediction method based on improved particle swarm optimization (IPSO). On the basis of traditional particle swarm optimization (PSO), chaos algorithm is added, and chaos algorithm is more random and more general and has deeper search ability to improve the overall prediction accuracy and prediction speed of the wavelet neural network. However, this method needs to continuously optimize the weights and parameters, and the model establishment is too complicated, so the speed slows down during the data training process. Jamal et al. (2020) and Khan et al. (2020) designed a multivariable phase space reconstruction Kalman filter method, which fully considered the coupling relationship between various energy sources, and used the five-step parameter trend method to dilute the influence of the old parameters on the current load forecasting. A larger training set was used to train the prediction models to improve the prediction accuracy. Because the loss function is set to a convex function, this does not guarantee that the global optimal solution can be achieved when the optimized nonconvex function is achieved. The final result will be greatly affected by the initial value of the parameter. It takes a very long time to calculate the loss function of all training data at the same time, so that the prediction accuracy may get a suboptimal solution. Matrenin et al. (2020) adopted the multitask learning method of deep structure (Pitchforth et al., 2021) to predict the complex energy load of the park type, using a combination of offline and online, but due to the fixed learning rate adopted by this network, it may cause network oscillations and make the speed of convergence slower and the prediction accuracy lower, and the optimal value is also not reached. Verma et al. (2021) solved the problem of overfitting and limited generalization ability of a single model, but they could not solve the application limitation of a single algorithm. The work of Wang et al. (2019) is composed of a variety of heterogeneous models, which overcome the shortcomings of the application limitation of a single algorithm, but the training set of each model is the same, and there are still problems of single model overfitting and limited generalization ability. Although, artificial intelligence, neural networks, support vector machines, and deep learning methods have made great progress of power system prediction. However, the abovementioned models have their specific application scopes, and they are less involved in the field of integrated energy.

On the basis of the abovementioned research, this paper proposes an adaptive IES load forecasting method based on the octopus model. First, this paper uses a generative adversarial network (GAN) to supplement the incomplete data in the integrated energy system, thereby reducing the data-induced inadequate forecasting accuracy. The problem is that the head mechanism of the octopus model is used to change the weight of the octopus' foot, according to the prediction accuracy to prevent local minimum points and improve the prediction accuracy. Then, the RMSProp algorithm is used to train the octopus model, and the adaptive learning rate is used to obtain the optimal parameters, which not only improves the model's adaptability to different types of data but also improves the prediction speed. Finally, it is based on the operation data onto the integrated energy system of a residential district in Shenyang. A simulation analysis of the algorithm proposed to this paper is carried out. The prediction results show that, in the comprehensive energy system load forecasting, the octopus model using RMSProp algorithm, XGBoost algorithm, and

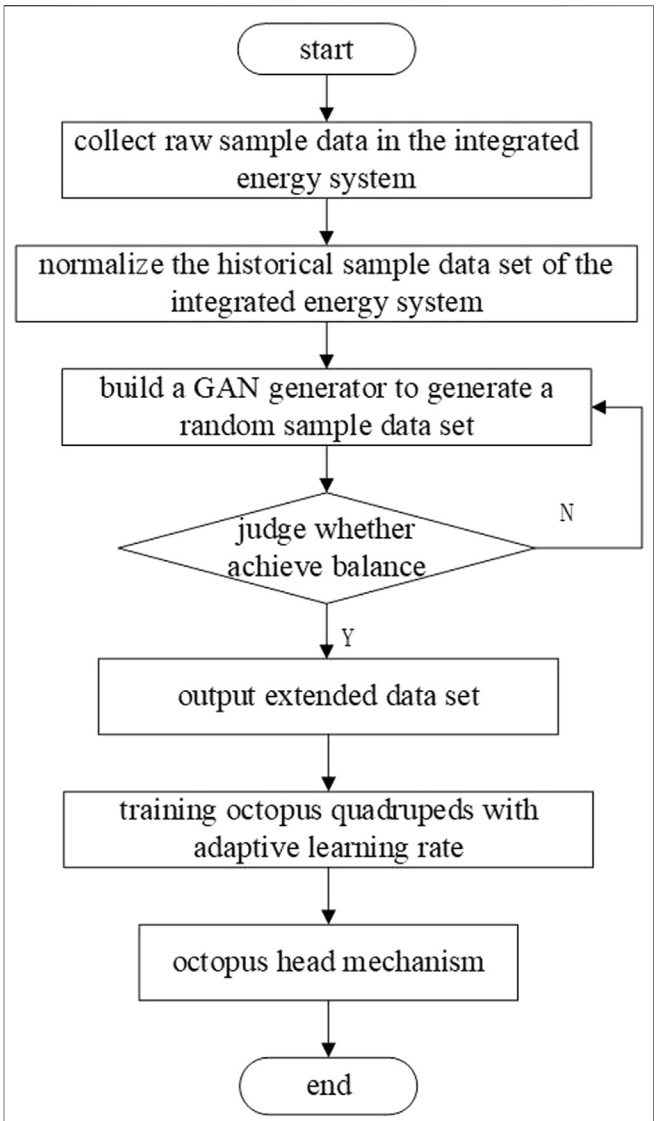

FIGURE 1 | Block diagram of the octopus model adaptive IES load forecasting method.

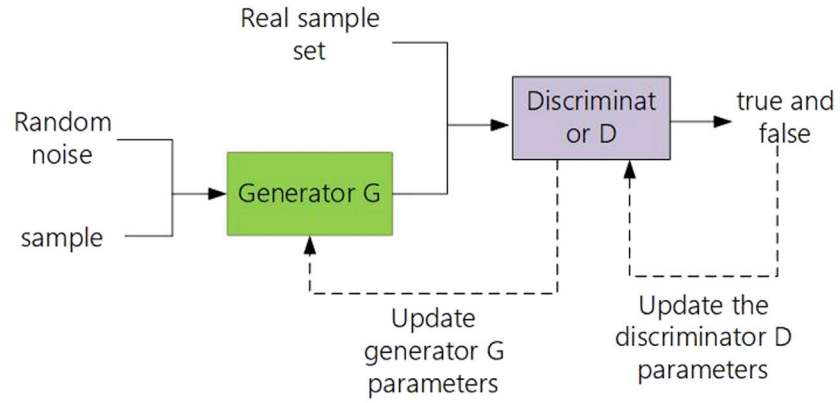

FIGURE 2 | Principle of the generative adversarial network.

GAN has good prediction accuracy and prediction speed and can better solve the problem that traditional neural network prediction models tend to fall into local optimal solutions and shortcomings such as limited application of predictive models.

OCTOPUS MODEL ADAPTIVE INTEGRATE ENERGY SYSTEM PREDICTION METHOD

This paper proposes an adaptive long and short-term IES load forecasting method based on the octopus model. First, we collect the original sample data and normalize the sample data. Then, the processed sample dataset is expanded by GAN. When the generator and the discriminator reach the Nash equilibrium, the expanded sample dataset can be generated from the incomplete sample dataset. We divide the extended sample dataset into six subdatasets according to the time dimension and ensure that each subdataset does not overlap with each other. Finally, we build an octopus model. Three of the six subdatasets are used as the training set, one is used as the test set, and the remaining 2 validation sets are input into the octopus model quadrupeds, and the octopus quadruped prediction results are integrated by weighted average through the octopus-head mechanism, so as to predict the electrical load, air load, and thermal load. The load forecasting model used in this paper can reduce the time for selecting the network model in the early stage and, at the same time, improve the accuracy of load forecasting. The overall idea of the forecasting method is shown in **Figure 1**.

GENERATIVE ADVERSARIAL NETWORK

Generative adversarial networks (GANs) consist of two parts, the generator and the discriminator. The generator uses the processing of random noise to generate pseudosamples, and the pseudosample value is similar to the real sample value. The discriminator compares the real sample value of the fake sample value generated by the generator, distinguishes the difference between the fake sample and the real value, and improves the recognition ability of its own network. The two realize the learning optimization processes through the game and finally reach a Nash equilibrium. The GAN schematic diagram is shown in **Figure 2**.

The generated data are generated by the original dataset training, so the generated data are similar to the original data. The data expanded by GAN will reduce the isolated points and discrete points of the original data. Using the data enhanced by GAN for load prediction will make the prediction accuracy higher. Since GAN already has good data generation capabilities, the use of extended data has little effect on the prediction results.

OCTOPUS FOUR-LEGGED MODEL

Long Short-Term Memory Principle

LSTM is a model that uses back propagation time to train a neural network. It is not a neuron, but a unit connected by layers. LSTM is a nonlinear prediction model, so it can build a larger and deeper recurrent neural network, effectively solving the problem of the disappearance of gradients in the prediction process, and is suitable for processing time series models. LSTM updates short-term memory through memory, forgets past memory information, and updates new information. But, when the sequence of continuous data becomes longer, the unfolding time step will be too long. The LSTM neural network has a long- and short-term memory structure, and LSTM is suitable for time series forecasting.

The memory unit memorizes the historical information about the sequence data together with the hidden state. The information about the memory unit is controlled by three gate units. The forget gate deletes the information about the memory unit according to h_{t-1} and x_t. The forget door is

$$f_t = \sigma\left(W_f\left[h_{t-1}\, x_t\right]\right) + b_f, \tag{1}$$

where $\sigma(\cdot)$ is the sigmoid activation function; $\boldsymbol{b}_f$ is the bias of the forgetting gate; and W_f is the weight of the forgetting gate.

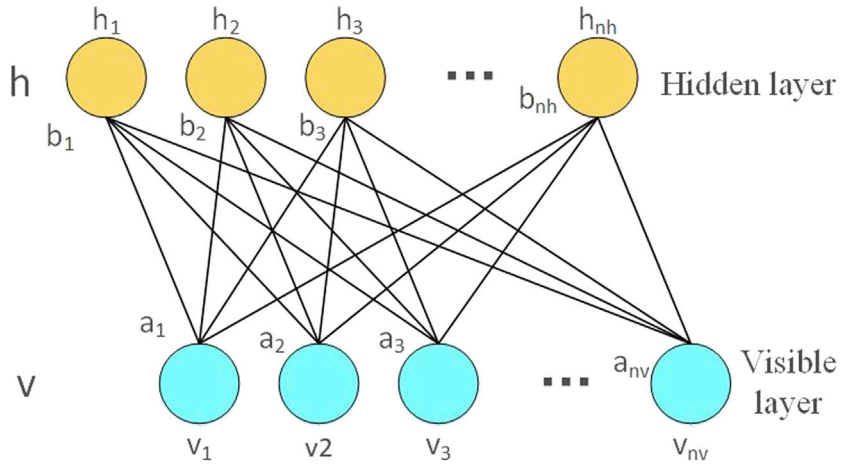

FIGURE 3 | RBM schematic.

The input gate adds information to the memory unit according to h_{t-1} and x_t, as shown in **Eqs 2** and **3**.

$$i_t = \sigma\left(W_i\left[h_{t-1}\ \ x_t\right]\right) + b_i, \tag{2}$$

$$\widehat{C}_t = \tanh\left(W_c\left[h_{t-1}\ x_t\right]\right) + b_c. \tag{3}$$

In the formula, i_t is the information that needs to be memorized; $\widehat{C}_t$ is the candidate memory unit, used to update the memory unit; W_i and W_c are the input gate weights; and b_i and b_c are the input gate biases.

After the calculation of the forget gate and the output gate is completed, the memory unit is updated using the following equation:

$$C_t = f_t \circ C_{t-1} + i_t \circ \widehat{C}_t, \tag{4}$$

where $\circ$ is the product of Hadamard.

The output gate determines h_t according to h_{t-1}, x_t, and C_t.

$$o_t = \sigma\left(W_0\left[h_{t-1}\ x_t\right]\right) + b_o, \tag{5}$$

$$h_t = o_t \circ \tanh C_t, \tag{6}$$

where W_o is the output gate weight; b_o is the output gate bias.

LSTM inherits the advantages of the recurrent neural network (RNN) well and has a long-term memory function. Compared with the prediction model constructed by the ordinary RNN, LSTM can solve the problems of gradient explosion and gradient disappearance. Therefore, LSTM has better performance in dealing with models that are highly correlated with time series.

Support Vector Machine Principle

SVM, as a relatively important learning method in machine learning, is based on statistical theory and supervised learning and can solve multivariate nonlinear problems well. The principle is to map the sample data one by one in a high-dimensional space, and this kind of mapping does not require a clear mapping function, so as to achieve a conversion from nonlinear to linear. Simply putting, it is to upgrade the data and linearize the data. Compared with other linear models, SVM can better solve the dimension problem with the premise of the same computational complexity. Therefore, the SVM model requires small storage space and strong algorithm robustness. SVM is proposed by a binary classification problem. When making predictions, a prediction curve is made through linear regression. SVM predicts data through regression fitting, and it does not have memory function. Therefore, SVM can reduce the impact of discrete points and isolated points on the prediction results.

Suppose the training set samples are $\{(x_i,y_i),\ i = 1,2,3,\ \ldots\ ,l\}$, where x_i is the input column vector of the *i*th training sample, $x_i = [x_i^1, x_i^2, \ldots, x_i^{d_1}]^T$, and $y_i \in R$ is the corresponding output value.

Suppose the linear regression function established in the high-dimensional feature space is

$$f(x) = w\Phi(x) + b. \tag{7}$$

Among them, $\Phi(x)$ is a nonlinear mapping function.

The ε linear insensitive loss function is defined as

$$\boldsymbol{L}(\boldsymbol{f}(\boldsymbol{x}), \boldsymbol{y}, \boldsymbol{\varepsilon}) = \begin{cases} 0 & |y - f(x)| \leqslant \varepsilon \\ |\boldsymbol{y} - \boldsymbol{f}(\boldsymbol{x}) - \boldsymbol{\varepsilon}| & |y - f(x)| > \varepsilon \end{cases}. \tag{8}$$

Among them, f(x) is the predicted value returned by the regression function, and y is the corresponding true value.

Restricted Boltzmann Machine Principle

Boltzmann machine (RBM) is a model with a two-layer neural network, which is a probability distribution model based on energy.

The Boltzmann machine is divided into a hidden layer h and a visible layer v (that is, the input layer and the output layer). The Boltzmann machine can be regarded as a fully connected graph; that is, each neuron is fully connected with all neurons in this layer and neurons in other layers. The principle of the RBM is shown in **Figure 3**. The RBM layer is not connected and is connected to all neurons in other layers, where v_i is the visible layer neuron, h_j is the hidden layer neuron, a_i is the visible layer, b_j

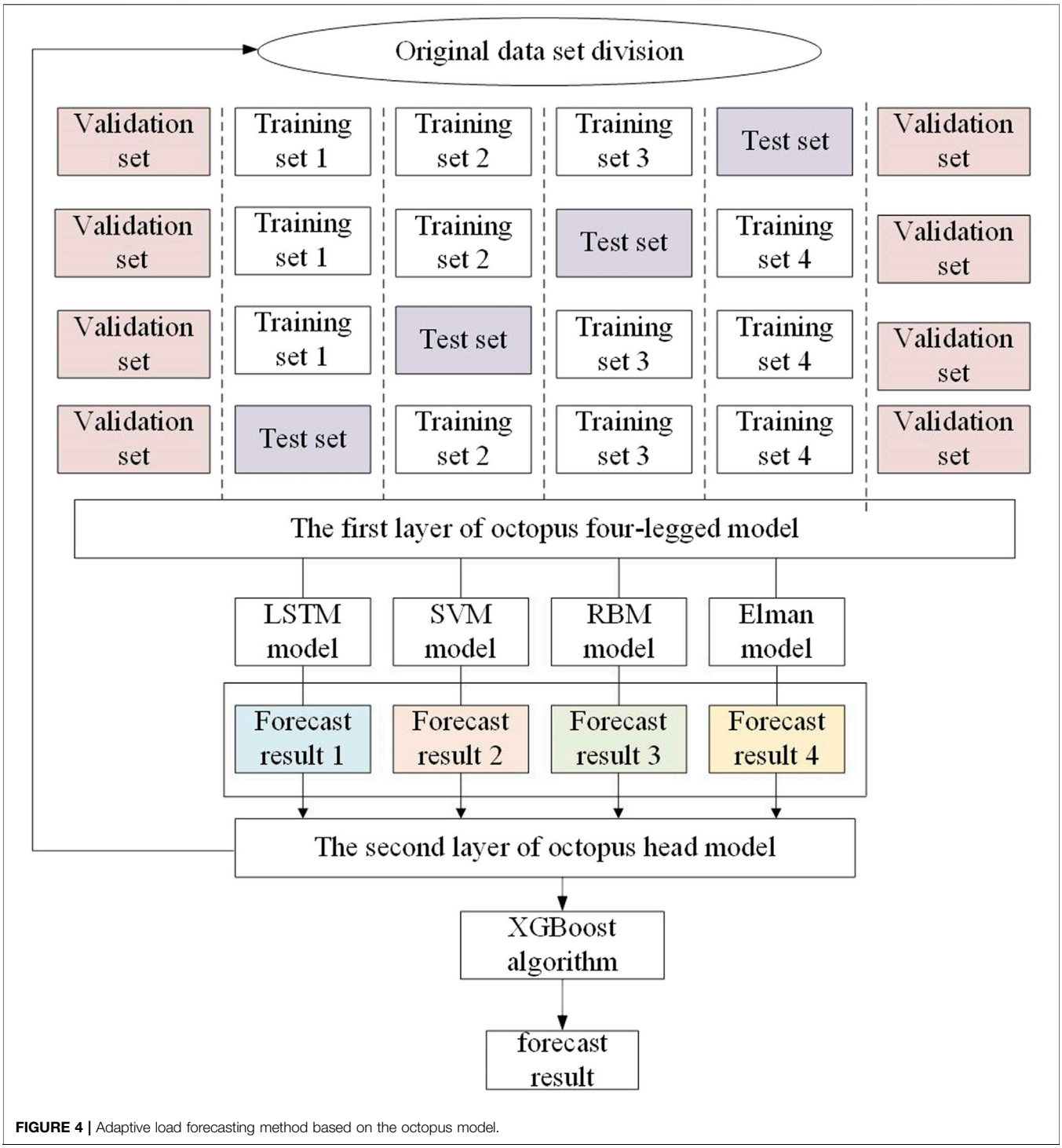

FIGURE 4 | Adaptive load forecasting method based on the octopus model.

is the hidden layer, and the W is the weight matrix. Its visible nodes are independent of other visible nodes, and there is no connection between hidden layer nodes. When the observation data are given in the visible layer, each node of the hidden layer is independent of each other. The RGM network is relatively simple and does not have a memory function. Important information entered early will be forgotten over time.

The hidden layer and visible layer of the RBM used in this paper are both binary, namely, $v_i \in \{0,1\}^n$, $h_j \in \{0,1\}^m$. The energy formula is shown as follows:

$$E(v,h) = -\sum_{i=1}^{n}\sum_{j=1}^{m} w_{ij} v_i h_j - \sum_{i=1}^{n} a_i v_i \sum_{j=1}^{m} b_j h_j. \quad (9)$$

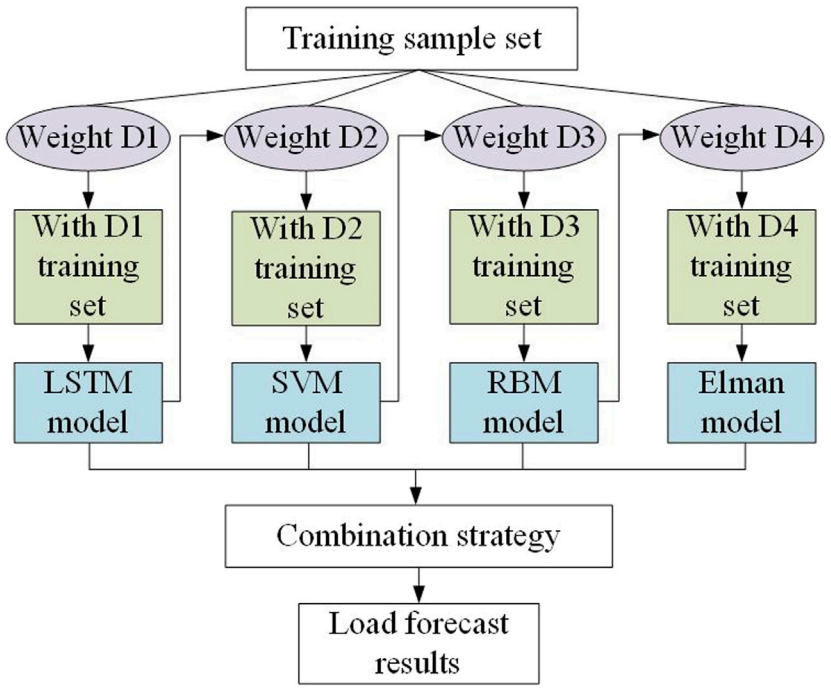

FIGURE 5 | Principle of the octopus-head model.

Among them, $w = [w_{ij}]^{n\times m}$, $a = (a_i)^n$, and $b = (b_j)^m$ constituted the parameter $\boldsymbol{\theta}$ of the RBM. m and n represent the number of visible units and winning units, v_i and h_j are the *i*th visible unit and the *j*th hidden unit, respectively, and w_{ij} is the weight between the visible unit and the hidden unit. a_i and b_j, respectively, represent the bias of the visible unit and the hidden unit.

The joint probability distribution p (v,h) of the visible unit and the hidden unit is defined as follows:

$$P(v,h) = \frac{1}{\sum_{v,h} e^{-E(v,h)}} e^{-E(v,h)}. \tag{10}$$

The corresponding two marginal probability distributions are

$$P(v) = \frac{1}{z}\sum_h e^{-E(r,h)}, \tag{11}$$

$$P(h) = \frac{1}{z}\sum_v e^{-E(v,h)}. \tag{12}$$

Elman Neural Network

The Elman neural network is a dynamic neural network. It adds a layer on the basis of the BP neural network, which is used as a delay factor. The Elman neural network can use, store, and feedback historical past time information. To a certain extent, the Elman neural network can perform load forecasting well. This paper uses a 4-layer Elman neural network. The connection to the output layer is similar to the feed forward network. The input layer unit only plays a role in signal transmission. The transfer function of the hidden layer unit adopts a nonlinear function. The receiving layer can be considered as a one-step delay operator. Elman is a dynamic network, adapts to time-varying characteristics, has a short-term memory function, can internally feedback, store, and use the output information of the past moment, and is better than the BP network in terms of computing power and network stability.

In this paper, the Elman neural network of particle swarm optimization (PSO) is used. PSO is an efficient and rapid optimization method, suitable for solving continuous weights in Elman. The position and speed update formulas are

$$V_i^{k+1} = \omega \cdot V_i^k + c_1 \cdot rand_1 \cdot (P_{best} - X_i) + c_2 \cdot rand_2 \cdot (G_{best} - X_i), \tag{13}$$

$$X_i^{k+1} = X_i^k + V_i^{k+1}, \tag{14}$$

where V_i^k and X_i^k are the velocity and position of particle *i* at the *k*th iteration, respectively. $rand_1$ and $rand_2$ are randomly generated in (-1,1), c_1 and c_2 are the learning factors, and ω is the particle weight coefficient. G_{best} is the global optimum, and P_{best} is the individual optimum.

ADAPTIVE LEARNING

The traditional training method can reach the global optimal solution only when the loss function is convex. For the concave function, it cannot ensure that the training of the neural network will definitely reach the global optimum. On the contrary, the loss function can easily reach the local optimum. At the same time, the traditional training process takes a long time, and the training result is greatly affected by the initial value.

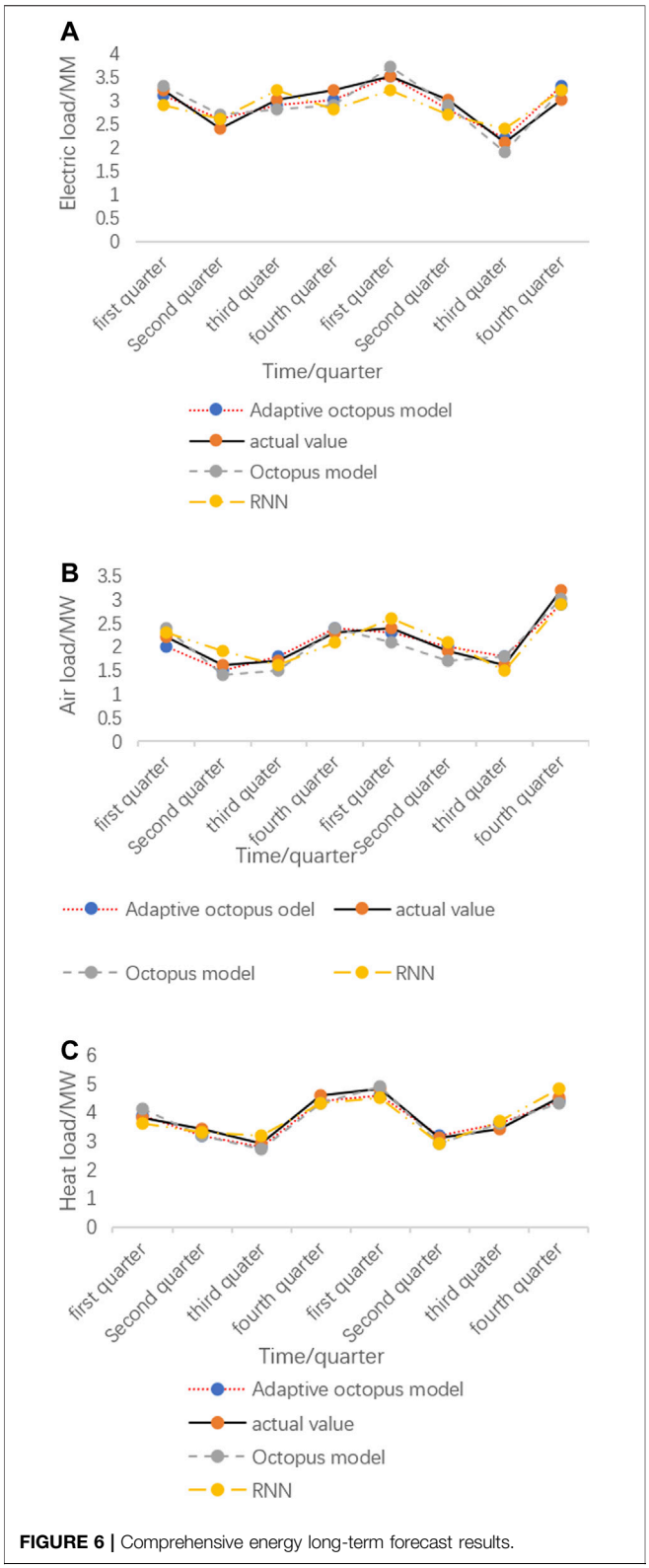

FIGURE 6 | Comprehensive energy long-term forecast results.

Aiming at the shortcomings of traditional training methods, the octopus model structure uses an adaptive learning rate algorithm to train four prediction models to obtain the best parameters of the octopus model. Taking into account that different parameters require different learning rates to be adjusted, if the learning rate is too large, it will cause the training of some parameters to miss the global optimum and reach the local optimum. If the learning rate is too small, the convergence speed of the parameters will be slow, and the training process will take a long time. Therefore, it is particularly important to set different learning rates for different parameters in training of the model.

The RMSProp optimization algorithm can improve the prediction accuracy of the octopus model and make the octopus model more adaptive. Compared with the AdaGrad optimization algorithm, the RMSProp algorithm uses a new exponential decay algorithm, which reduces the impact on historical data. At the same time, RMSProp introduces a new parameter ρ, which can be expressed as the second derivative of the gradient value, which is used to control the decay rate of the historical gradient value. Therefore, RMSProp algorithm has better adaptability than AdaGrad. The main body of the algorithm executes the following loop steps and will not stop until the stop condition is reached.

We take out the small batch of data $\{x_1, x_2, \ldots, x_m\}$. The target corresponding to the data is denoted by y_i.

We calculate the gradient based on the small batch data according to the following formula:

$$g = \frac{1}{m}\nabla_\omega \sum_i L\left(f\left(x_i;\omega\right), y_i\right). \tag{15}$$

We accumulate the square gradient and refresh r, and the process is as follows:

$$r = \rho r + \left(1-\rho\right) g \odot g. \tag{16}$$

We calculate the parameter update amount as follows:

$$\Delta\omega = -\frac{\sigma}{\sqrt{\delta + r}} \odot g. \tag{17}$$

We update parameters according to Δω:

$$\omega = \omega + \Delta\omega. \tag{18}$$

The traditional stochastic gradient descent maintains a single learning rate to update all weights, and the learning rate does not change from the training process. The RMSProp algorithm uses different learning rates when optimizing deep neural networks, which is efficient and practical.

OCTOPUS MODEL

The octopus-head model in **Figure 4** uses XGBoost, which is an optimized integrated tree model, improved and extended for the gradient boosting tree model. The integrated model of the tree is as follows:

$$\bar{y}_i = \sum_{k=1}^{K} f_k\left(x_i\right), f_k \in F, \tag{19}$$

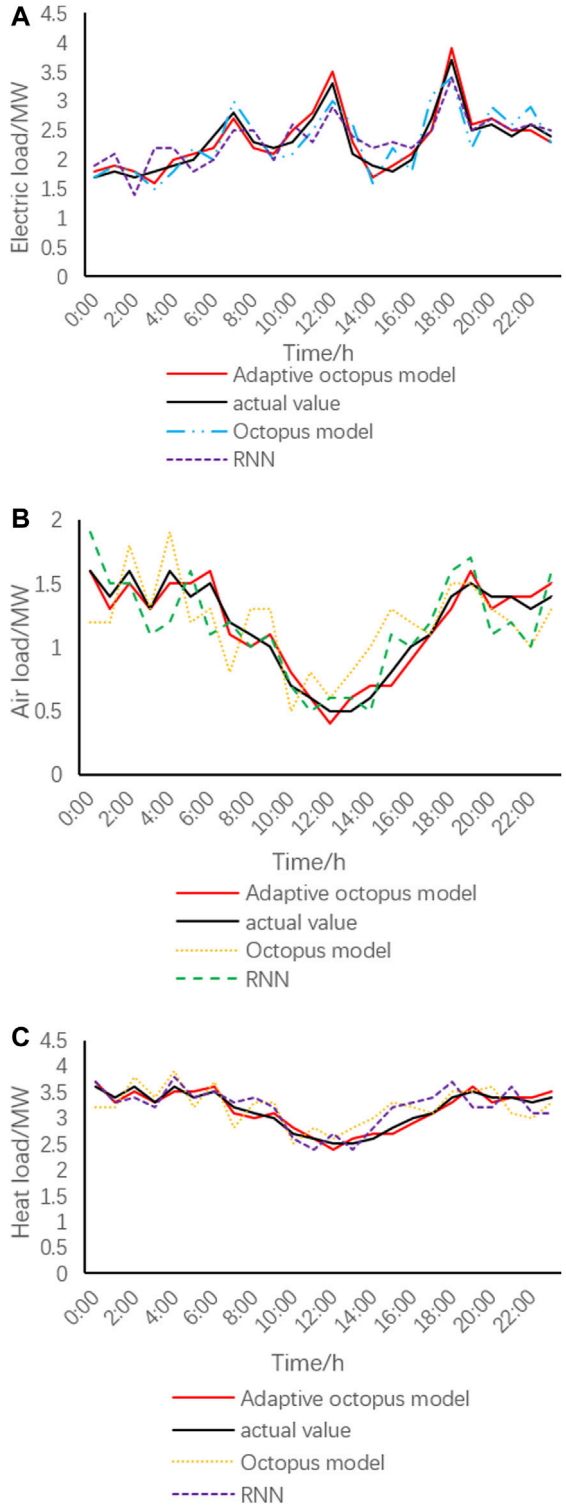

FIGURE 7 | Comprehensive energy short-term daily load forecast results.

where $\overline{y}_i$ is the predicted value of the *i*th sample; x_i is the feature vector of the *i*th data point, f_k is the structure *q* of the *k*th tree and the leaf weight w, K is the number of trees, and F is the collection space of trees.

The loss function of the model can be expressed as

$$L = \sum_{i=1}^{n} l(y_i, \overline{y}_i) + \sum_{k=1}^{K} \varepsilon(f_k). \tag{20}$$

In the formula, the first half is the error between the predicted value and the true value, and the second half is the complexity of the tree.

In the process of minimizing the sequence, the loss function is reduced by adding the increment function f_i (x_i). The objective function of the *m*th round is

$$\begin{aligned} L^{(m)} &= \sum_{i=1}^{n} l(y_i, \overline{y}_i) + \sum_{k=1}^{K} \varepsilon(f_k) \\ &= \sum_{i=1}^{n} l\left(y_i, \overline{y}_i^{(m-1)} + f_m(x_i)\right) + \varepsilon(f_m). \end{aligned} \tag{21}$$

For **Eq. 21**, the second-order Taylor expansion is used to approximate the objective function. This results in

$$\begin{aligned} L^{(m)} &= \sum_{j=1}^{T} \left[\left(\sum_{i \in I_j} g_i \right) w_j + \frac{1}{2} \left(\sum_{i \in I_j} h_i + \lambda \right) w_j^2 \right] + \mu T, \\ g_i &= l'\left(y_i, \hat{y}_i^{(t-1)}\right), \\ h_i &= l''\left(y_i, \hat{y}_i^{(t-1)}\right). \end{aligned} \tag{22}$$

We find the partial derivative of w to get

$$w_j = -\frac{G_j}{H_{j+\lambda}}. \tag{23}$$

Substituting the weights into the objective function, we can get

$$L^{(m)} \cong -\frac{1}{2} \sum_{j=1}^{T} \frac{G_j}{H_{j+\lambda}} \mu T. \tag{24}$$

The octopus-head model mainly uses the weighted average method to train the octopus' quadrupeds with initial weights. The head adopts an integrated learning method to train the weight parameters according to the prediction accuracy of the four prediction models. The weights are changed according to the prediction accuracy of the octopus' quadrupeds, so that the weight of high prediction accuracy increases, and the weight of the other octopus' feet with relatively low prediction accuracy decreases, and then, the octopus' quadruped after weight adjustment continues to train and repeats this until the specified number of times is reached. Finally, these prediction accuracies will be integrated according to the weight, and the final prediction result will be obtained.

CASE ANALYSIS

Screening of Related Factors

We use the influencing factors and load data onto m days before the forecast date to predict the comprehensive energy load data onto the forecast day, and the model output is the comprehensive energy data onto the forecast day, as shown in **Eq. 20**.

$$\widehat{L}_d = \left\{\widehat{l}_{0,d})\widehat{l}_{1,d})\cdots)\widehat{l}_{t,d}(\cdots)\widehat{l}_{T,d}\right\}. \quad (25)$$

In the formula, T is the number of points to be predicted.

The comprehensive energy historical load data are shown in **Eq. 21**.

$$L_{t,d-w} = \{l_{t,d-1})l_{t,d-2})\cdots)l_{t,d-m}\}. \quad (26)$$

In the formula, $l_{t,d-m}$ is the load at time t on m days before the forecast.

The comprehensive energy load influencing factors are

$$F_{t,d} = \{w_{t)d}(e_{t)d})\cdots)h_{t,d}\}. \quad (27)$$

Data Normalization

Formula **(23)** is used to normalize the comprehensive energy data and obtain real sample data of temperature, humidity, date, and economy.

$$x_{std} = \frac{x - x_{min}}{xmin_{max} + 1}. \quad (28)$$

In the formula, x represents each sample data of the integrated energy system, x_{max} represents the sample data with the largest absolute value in each sample data set, x_{min} represents the sample data with the smallest absolute value, and x_{std} represents the normalized value.

Result Analysis

This paper selects the comprehensive energy system operation data of a residential district in Shenyang from January 1, 2009, to January 1, 2020, and the comprehensive energy operation data of the district in the first week of August 2019 for analysis. The comprehensive energy data from 2009 to 2018 are used as the training set quarterly to predict the comprehensive energy operation data in 2019 and 2020; the data from the first 8 days of August 2019 are used as the training set to predict the electrical load, air load, and heat load data.

In order to verify the effect of adaptive load forecasting based on the octopus model, the long-term and short-term forecasting were carried out separately, and two cases were set up.

Case 1: considering the long-term forecast of the coupling of electricity, gas, and heat loads, the octopus model is used to analyze the comprehensive energy data from 2009 to 2018 to predict the electricity, gas, and heat loads in 2019 and 2020.

Case 2: considering the short-term forecast of the coupling of electricity, gas, and heat loads, the octopus model is used to predict one day's load data.

The long-term and short-term results of comprehensive energy in the two cases are shown in **Figure 5** and **Figure 6**, respectively.

It can be seen that whether it is a long-term forecast or a short-term forecast of electricity, gas, and heat load, the forecast curve has good tracking ability. Only when the real value fluctuates greatly, the predicted value will have a large error. From **Figure 7**, it can be seen that the electrical load fluctuates greatly throughout the year and the heat load changes significantly with the seasons.

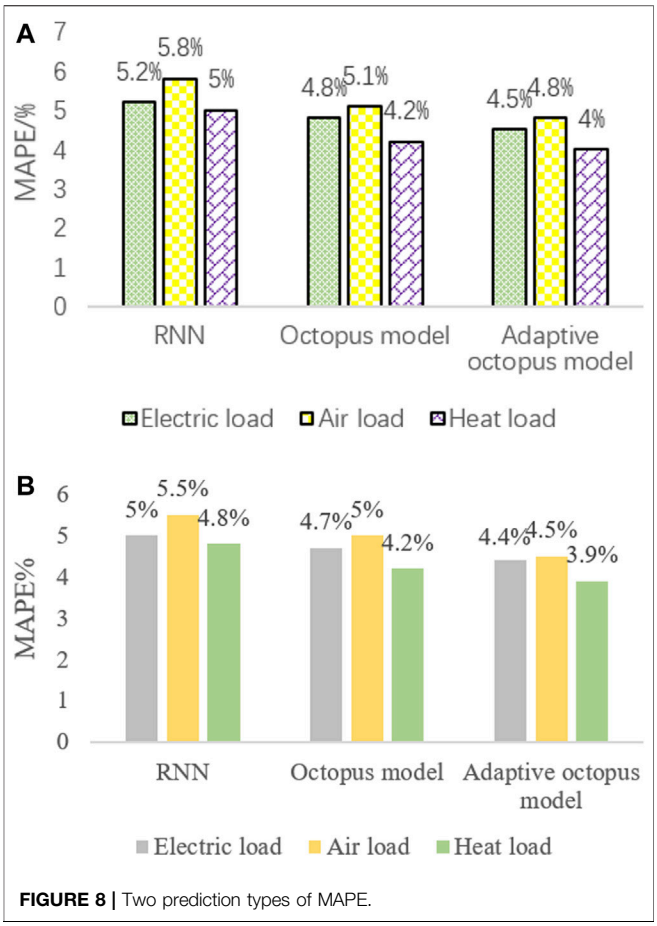

FIGURE 8 | Two prediction types of MAPE.

It can be seen from **Figure 5** that the electricity load fluctuates greatly throughout the day, with little electricity consumption during the day and large electricity consumption at night.

This article uses MAPE as the error evaluation standard, and the calculation formula of MAPE is

$$\delta_{MAPE} = \frac{1}{N}\sum_{t=1}^{N}\left|\frac{\overline{y}_i - y_i}{y_i}\right| \times 100\%. \quad (29)$$

In the formula, y_t is the actual value; $y_t^{'}$ is the predicted value; and N is the number of predicted points.

Figure 8 shows the comparison between the MAPE value predicted by the octopus model and the MAPE value predicted by the ordinary RNN in two cases. The MAPE value of the adaptive IES load forecasting method is between 4 and 6%. Both of the MAPE values are not large.

It can be seen from **Figure 8** that the accuracy of short-term forecasts is higher than that of long-term forecasts. In the long-term and short-term forecasting, the adaptive octopus model has higher prediction accuracy and smaller error than the traditional load forecasting and ordinary octopus model and has a small MAPE value.

Compared with the results of gas and heat load forecasting, heat load forecasting has higher forecast accuracy, which is mainly due to the fact that heat load is greatly affected by temperature, humidity, economy, and date changes, especially

with a strong relationship between temperature and relevance. However, there are many uncontrollable factors of electricity and gas load, so the prediction accuracy is lower than that of heat load. From the perspective of long-term and short-term prediction results, the octopus model has better short-term prediction results. This is mainly due to the large training sample dataset for short-term load prediction and the small training error of the model.

CONCLUSION

Aiming at the background of integrated energy system, this paper proposes an adaptive integrated energy load forecasting method based on octopus model. This method not only uses the octopus model to effectively reduce the risk of model overfitting and prevents local minima but also improves the convergence speed of the model through the RMSProp algorithm and improves the prediction accuracy. It has high application value in the multielement load forecasting of the integrated energy system.

With the development of the energy Internet, the integrated energy system will receive more and more attention. The adaptive IES load forecasting method based on the octopus model will have a higher development and application in the energy Internet system. The model's adaptive IES load forecast method has extremely high forecasting accuracy in ultra-short-term forecasting. In the future, it is hoped that, through further applications, it can predict abnormal conditions more accurately and improve the forecasting results.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/Supplementary Material; further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

NZ conceived the idea for the manuscript and wrote the manuscript with input from XP, YW, MZ, MC, and WS. All authors have read and agreed to the published version of the manuscript.

REFERENCES

AlDahoul, N., Essam, Y., Kumar, P., Ahmed, A. N., Sherif, M., Sefelnasr, A., et al. (2021). Suspended Sediment Load Prediction Using Long Short-Term Memory Neural Network. *Sci. Rep.* 11 (1), 7826. doi:10.1038/s41598-021-87415-4

Bian, H., Zhong, Y., Sun, J., and Shi, F. (2020). Study on Power Consumption Load Forecast Based on K-Means Clustering and FCM-BP Model. *Energ. Rep.* 6 (S9), 693–700. doi:10.1016/j.egyr.2020.11.148

Dai, Y., and Zhao, P. (2020)., 279. A Hybrid Load Forecasting Model Based on Support Vector Machine with Intelligent Methods for Feature Selection and Parameter Optimization. *Appl. Energ.*, 115332. doi:10.1016/j.apenergy.2020.115332

Daniel, R., Teixeira, B., Pedro, F., Gomes, L., Omid, A., and Zita, V. (2020). Using Diverse Sensors in Load Forecasting in an Office Building to Support Energy Management[J]. *Energ. Rep.* 6 (S8). doi:10.1016/j.egyr.2020.11.100

Dong, Y., Dong, Z., Zhao, T., Li, Z., and Ding, Z. (2021). Short Term Load Forecasting with Markovian Switching Distributed Deep Belief Networks. *Int. J. Electr. Power Energ. Syst.* 130, 106942. doi:10.1016/j.ijepes.2021.106942

Guo, X., Gao, Y., Li, Y., Zheng, D., and Shan, D. (2021). Short-term Household Load Forecasting Based on Long- and Short-Term Time-Series Network. *Energ. Rep.* 7 (S1), 58–64. doi:10.1016/j.egyr.2021.02.023

Guo, X., Zhao, Q., Zheng, D., Ning, Y., and Gao, Y. (2020). A Short-Term Load Forecasting Model of Multi-Scale CNN-LSTM Hybrid Neural Network Considering the Real-Time Electricity price. *Energ. Rep.* 6 (S9), 1046–1053. doi:10.1016/j.egyr.2020.11.078

Jamal, F., Hamed, H-D., and Abbas, K. (2020).Multi-year Load Growth-Based Optimal Planning of Grid-Connected Microgrid Considering Long-Term Load Demand Forecasting: A Case Study of Tehran, Iran[J]. *Sustainable Energ. Tech. Assessments* 42. doi:10.1016/j.seta.2020.100827

Khan, Z. A. A., Ullah, A., Ullah, W., Rho, S., Lee, M., and Wook, S. W. B. (2020). Electrical Energy Prediction in Residential Buildings for Short-Term Horizons Using Hybrid Deep Learning Strategy. *Appl. Sci.* 10 (23), 8634. doi:10.3390/app10238634

Matrenin, P. V., Manusov, V. Z., Khalyasmaa, A. I., Antonenkov, D. V., Eroshenko, S. A., and Butusov, D. N. (2020). Improving Accuracy and Generalization Performance of Small-Size Recurrent Neural Networks Applied to Short-Term Load Forecasting. *Mathematics* 8 (12), 2169. doi:10.3390/math8122169

Pitchforth, D. J., Rogers, T. J., Tygesen, U. T., and Cross, E. J. (2021). Grey-box Models for Wave Loading Prediction. *Mech. Syst. Signal Process.* 159, 107741. doi:10.1016/j.ymssp.2021.107741

Verma, S., and Bala, A. (2021). Auto-scaling Techniques for IoT-Based Cloud Applications: a Review. *Cluster Comput.* doi:10.1007/s10586-021-03265-9

Wang, R., Sun, Q., Ma, D., and Liu, Z. (2019). The Small-Signal Stability Analysis of the Droop-Controlled Converter in Electromagnetic Timescale. *IEEE Trans. Sustain. Energ.* 10 (3), 1459–1469. doi:10.1109/tste.2019.2894633

Conflict of Interest: NZ is employed by the Economic and Technological Research Institute of State Grid Liaoning Electric Power Co., Ltd. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Energy Research

ORIGINAL RESEARCH
published: 07 February 2022
doi: 10.3389/fenrg.2021.814489

Adaptive Droop Control of the VSC-MTDC Distribution Network Considering Power–Voltage Deviation

Yang Li[1], Jianjun Zhao[1], Huan Liu[1], Qiankun Kong[1], Yanhui Zhao[2], Long Cheng[2] and Zhenhao Wang[2]*

[1]Smart Distribution Network Center, State Grid Jibei Electric Power Co., Ltd., Qinhuangdao, China, [2]Key Laboratory of Modern Power System Simulation and Control and Renewable Energy Technology, Ministry of Education (Northeast Electric Power University), Jilin, China

In order to realize the unbalanced power optimally allocated and the DC voltage stably controlled after disturbance, an adaptive droop control method considering power and voltage deviation is proposed based on the traditional voltage–power droop control of a voltage source converter-based multi-terminal direct current (VSC-MTDC) distribution network. The inherent constraint that the unbalanced power is proportionally distributed according to its capacity under the traditional droop control is broken in the proposed method to realize the reasonable transfer of unbalanced power and to reduce the overload risk of smaller capacity VSCs; the "dead zone" is appropriately set to relax the operating range of the VSC to a certain extent by a power deviation factor being introduced in the droop characteristic curve. The corresponding MATLAB/Simulink simulation model of the five-terminal DC power distribution network is established and compared with the electromagnetic transient model under the traditional droop control. Finally, the simulation results verify the effectiveness and control effects of the proposed control method.

Keywords: VSC-MTDC distribution network, adaptive droop control, power–voltage deviation, unbalanced power, DC bus voltage

OPEN ACCESS

Edited by:
Qiuye Sun,
Northeastern University, China

Reviewed by:
Yizhen Wang,
Tianjin University, China
Xuguang Hu,
Northeastern University, China

****Correspondence:***
Yanhui Zhao
1932644637@qq.com

Specialty section:
This article was submitted to Smart Grids, a section of the journal Frontiers in Energy Research

Received: *13 November 2021*
Accepted: *31 December 2021*
Published: *07 February 2022*

Citation:
Li Y, Zhao J, Liu H, Kong Q, Zhao Y, Cheng L and Wang Z (2022) Adaptive Droop Control of the VSC-MTDC Distribution Network Considering Power–Voltage Deviation. Front. Energy Res. 9:814489. doi: 10.3389/fenrg.2021.814489

INTRODUCTION

The continuous maturity of flexible DC equipment and control technologies in the field of electricity transmission has greatly promoted the development of DC power distribution (Li and Lao, 2017; Liao et al., 2018). Compared with the traditional AC power distribution network, the DC power distribution network has many advantages, such as lower loss, larger transmission capacity, higher power quality and power supply reliability, and easier power control, regardless of the frequency and voltage phase, more convenient for large-scale access to clean energy and lower environmental pollution (Yang et al., .2015; Jaynendra et al., 2019; Li et al., 2021a), and can effectively isolate AC side faults and disturbances in parallel with the AC system (Xu et al., 2019; Zhao et al., 2019). As an important basis of the energy internet and smart grids, a reliable, flexible, and efficient flexible DC power distribution network has gradually become an important guarantee for the safe and economic operation of the power system and power supply at a high service level (Gao et al., 2019). Therefore, the construction and development of the flexible DC power distribution network is of great significance to meet the needs of energy conservation, emission reduction, and comprehensive energy utilization in various countries, to improve the intelligent level of power supply, to promote

the transition from traditional power grids to the energy internet, and to build a green and environmentally friendly energy society (Li et al., 2021b; Li et al., 2022).

The three-level VSC-MTDC power distribution network has the characteristics of multi-source power transmission, multi-drop power reception, and system power flow flexibly regulation and control and has become an effective solution to develop and reform the power supply mode in the future (Li et al., 2019). The power flow of the DC power distribution network has frequent fluctuations, and the transient process is very short, so it brings great challenges to the coordination of VSCs, the power optimal dispatching, and the voltage stability control (Beerten and Belmans, 2013; Wang and Barnes, 2014). Therefore, as a typical multi-point control, the droop control has become the hot point of the current research for the fast response capability to the change of power flow. In the aspect of the control strategy, Pedram and Mohsen, 2018 proposed a distributed control method of the DC system based on the main controller and low-bandwidth communication and realized accurate power allocation by setting droop gain, but this method depends on the communication between VSCs to a certain extent. Chen et al., 2018 proposed an adaptive droop control method for the multi-terminal DC system based on the compensation governor with synthetically considering the dynamic voltage and power deviation of the DC network, which improved the system steady-state characteristics and dynamic response. Wang et al., 2019a addressed the problem that a fixed droop control coefficient will reduce the DC voltage control capability of the entire MTDC system and proposed an adaptive droop control scheme based on the DC voltage deviation factor and power distribution factor to ensure that the MTDC system maintains a high power sharing capability. Wang et al., 2020a derived the VDM model related to the DC voltage through the VSC-MTDC generalized linear model and proposed a droop coefficient adaptive method, which can realize the effective control of the system DC voltage. Wang et al., 2020b proposed a structure-changed master–slave control method based on the equal load rate based on the master–slave control method of DC distribution systems, which can reduce the DC voltage deviation when disturbance occurs. In Qusay and Xie, 2018, a transformer less H-bridge inverter with a series power flow controller is designed to control the transmission power of PCC, and its power supply connection interface adopted the U–P droop control strategy, which improved the control flexibility of the system, but the DC transient overvoltage is high in the process of fluctuation. Li et al., 2017 basically realized the reasonable distribution of active power and the stable control of AC side voltage of each VSC according to the unified adaptive droop control based on dynamic reactive power limiter, but the DC side voltage of VSC had not been deeply analyzed and verified by simulation. In the aspect of model analysis, Rouzbehi et al., 2014 realized the economic operation of the DC system by an improved optimal power flow algorithm, but to some extent, this way of accurately controlling voltage and power by modifying the droop coefficient accordance with power flow optimization results reduced the response speed of droop control to power flow change. Han et al., 2016 proposed a hybrid MTDC system decentralized autonomous control based on the consensus algorithm considering actual requirements of wind power grid connection and power transmission; the model convergence performed well, and the global information acquired fast under power fluctuation. The above documents were the necessary combination and improvement of the traditional droop control at different angles, which improved the distribution accuracy of the active power assumed by each VSC, but none of them really realized the isochronous control of DC voltage.

In terms of VSC-MTDC system stability modeling, Wang et al., 2019b proposed a construction method of the characteristic equation for the microgrid system composed of phase-locked loop DG. Compared with the traditional state-space matrix research method, this method can determine the phase angle margin and stability margin of the system stability, and the Routh criterion can be used to simply judge the stability. In reference to the independent power supply system composed of multiple batteries, Wang et al., 2021a constructed a forbidden zone criterion based on the regression ratio matrix accordingly to establish a state matrix and a rate of return matrix and proposed a sag coefficient stable area analysis method. Wang et al., 2021b, on the basis of Wang et al., 2021a, proposed a reduced-order aggregation model based on the Routh criterion and the balanced truncation method, which can solve the problem of large input–output mapping errors between the original system and the reduced-order system. Ma et al., 2021 proposed a dual-predictive control method based on adaptive error correction (DPCEC) applied to FW-VSIs for AC microgrids, which can deal with and correct the influence of different negative factors and realize the voltage source inverter real-time tracking of the reference value and the accurate value.

At present, the research on the control strategies of the MTDC power network mostly focuses on the transmission network, and the load fluctuation and power flow change of the distribution network are more complex, which puts forward higher requirements for the design of the control system.

On the basis of traditional droop control, an adaptive droop control of the VSC-MTDC distribution network considering power–voltage deviation is proposed in this article, and the electromagnetic transient model of the five-terminal VSC-MTDC distribution network based on the MATLAB/Simulink platform is built to verify the effectiveness of the proposed adaptive control strategy by simulation according to different system operating conditions. The main contributions of this article are as follows:

1) The proposed control strategy can break the traditional droop control restriction of the active power distribution according to a fixed ratio under system disturbance; it can adaptively adjust the droop coefficient to realize the optimal distribution of power and effectively prevent the overload of the smaller-capacity converters.
2) By superimposing the constant voltage control link in the improved adaptive droop control strategy, the system voltage stability before and after the transient process can be effectively guaranteed, and the error adjustment can be realized without relying on communication.

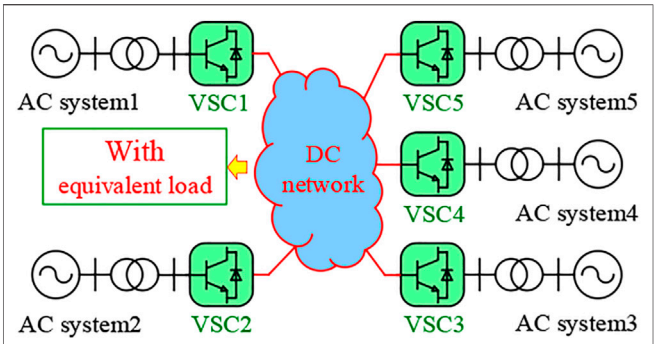

FIGURE 1 | Typical structure of five-terminal DC power distribution-based VSC.

TYPICAL STRUCTURE OF THE VSC-MTDC POWER DISTRIBUTION

The typical structure of the VSC-MTDC distribution network is shown in **Figure 1**. Taking the five-terminal power system as an example, the AC system is connected to the DC network with equivalent load through the corresponding VSC. One of the VSCs is set up as the main station and adopts the constant DC voltage control to maintain the DC bus voltage stability. The other four VSCs are slave converters, which adopt the adaptive droop control strategy considering the power–voltage deviation to realize the system power optimal distribution and to ensure the stable operation of the DC system according to the requirement of VSCs, the equipment connected to the AC side, the topology of the VSC-MTDC system, and the dispatching plan. The following will carry on the detailed analysis to the VSCs which adopt the droop control.

ADAPTIVE DROOP CONTROL OF VSC-MTDC

The adaptive control of the VSC-MTDC distribution network requires that each VSC can make independent decisions and update the decision value in real time. When the loads, power flow direction, and grid structure of the network change, each VSC controller should be able to maintain the system stable and reliable operation between the allowable power and voltage regions.

DC Voltage Droop Characteristic Analysis

For the traditional voltage droop control, set the positive direction as the absorption power of VSC, so the relationship between DC voltage U_{DCi} and output current I_{DCi} can be expressed as follows

$$I_{DCi}^{ref} - I_{DCi} + K_{droopi}^{0}\left(U_{DC}^{ref} - U_{DCi}\right) = 0, \tag{1}$$

where U_{DC}^{ref} is the DC side voltage reference value of VSC; I_{DCi}^{ref} is the internal loop current reference value of VSC$_i$; and K_{droopi}^{0} is the droop coefficient defined by the U–I relationship, $K_{droopi}^{0} > 0$. Also, $P_{DCi} = U_{DCi}I_{DCi}$, so the output power of the VSC is

$$P_{DCi} = -K_{droopi}^{0}U_{DCi}^{2} + \left(K_{droopi}^{0}U_{DC}^{ref} + I_{DCi}^{ref}\right)U_{DCi}. \tag{2}$$

The U–P characteristic curve of the VSC drawn by **Eq. 2** is shown in **Figure 2**. As can be seen from **Figure 2**, the operating characteristic curve of the VSC is a parabola opening to the left (only taking the upper half of the symmetrical axis according to the physical meaning). The limit operating point $M(P_{iMAX}, U_{iMIN})$, respectively, corresponds to the power maximum value and the voltage minimum value, and there is

$$\begin{cases} P_{iMAX} = \dfrac{1}{4}K_{droopi}^{0}U_{DC}^{2ref} + \dfrac{1}{2}P_{DCi}^{ref} + \dfrac{I_{DCi}^{2ref}}{4K_{droopi}^{0}} \\ U_{iMIN} = \dfrac{1}{2}U_{DC}^{ref} + \dfrac{I_{DCi}^{ref}}{2K_{droopi}^{0}}. \end{cases} \tag{3}$$

where P_{DCi}^{ref} is the output power reference value of VSC, $P_{DCi}^{ref} = U_{DC}^{ref}I_{DCi}^{ref}$; P_{DCHi} and P_{DCLi} are the upper and lower limits of the operating power of VSCs, respectively; and U_{DCHi} and U_{DCLi} are the upper and lower limits of the DC side voltage of VSCs, respectively. The tangent point $N(P_{DCi}^{ref}, U_{DC}^{ref})$ of the operation characteristic curve and the drooping characteristic curve 1 is the optimal operation state point of VSC.

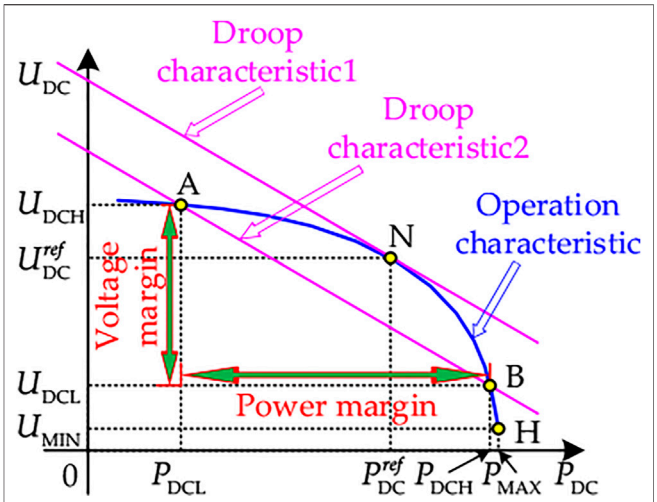

FIGURE 2 | Characteristic curves of *U–P* operating.

From **Eq. 2** and **Eq. 3**, it can be seen that the DC voltage regulation and power allocation of the VSC-MTDC power distribution network are determined by the droop coefficient. The selection of its value affects the dynamic performance and stability of the whole VSC-MTDC distribution network, so it is necessary to optimize the droop characteristic curve according to the characteristics of the power node (VSC) and the DC network.

At the same time, the voltage safety margin and power security margin should be considered in the operation of the VSC-MTDC distribution network (i.e., the AB section of the operating characteristic curve in **Figure 2**), which not only satisfies the

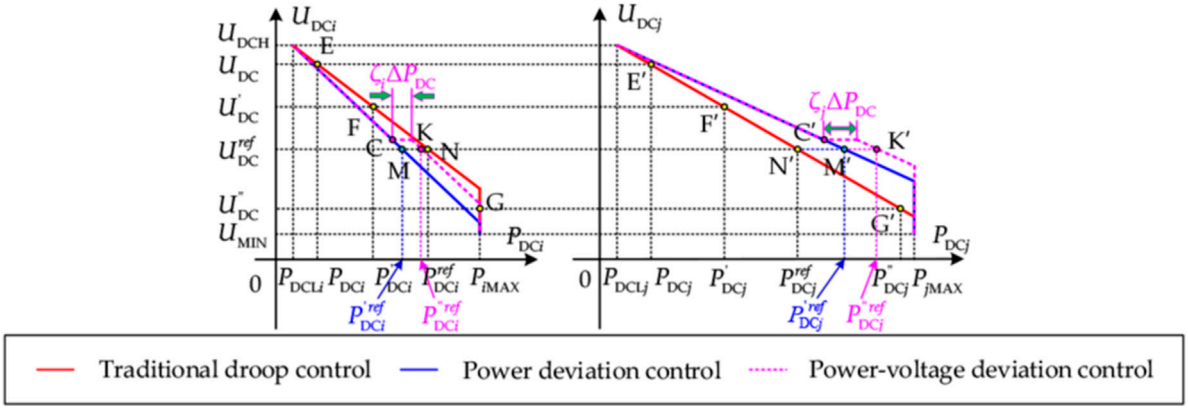

FIGURE 3 | Characteristic curves of *U–P* droop control.

power balance equation of the DC network but also satisfies the fixed boundary conditions of DC nodes' voltage amplitude and VSCs' operating power, that is,

$$P_{\mathrm{DC}i} - U_{\mathrm{DC}i} \sum_{j \in i} Y_{ij} U_{\mathrm{DC}j} = 0, \tag{4}$$

$$\begin{cases} U_{\mathrm{DCL}i} \le U_{\mathrm{DC}i} \le U_{\mathrm{DCH}i} \\ P_{\mathrm{DCL}i} \le P_{\mathrm{DC}i} \le P_{\mathrm{DCH}i}. \end{cases} \tag{5}$$

where $j \in i$ represents the node connected to node i.

Droop Characteristic Optimization of VSC Considering Power–Voltage Deviation

According to the droop characteristic of the VSC-MTDC power distribution network, the output error signal χ of the controller is set as follows:

$$\chi = P_{\mathrm{DC}i}^{ref} - P_{\mathrm{DC}i} + K_{\mathrm{droop}i}\left(U_{\mathrm{DC}}^{ref} - U_{\mathrm{DC}i}\right). \tag{6}$$

At steady-state operation, the error signal output by the VSC controller is 0 (that is $\chi = 0$). We respectively set the upper limit of operating power of the VSC$_i$ and VSC$_j$ in the network as $P_{i\mathrm{MAX}}$ and $P_{j\mathrm{MAX}}$; meanwhile, there is $P_{i\mathrm{MAX}} < P_{j\mathrm{MAX}}$. In the case of ignoring the DC line resistance, it can be considered that the power loss of the DC network is 0 and the voltage drop is 0. When disturbance occurs, the stable operating point of the VSC$_i$ changes from $E(P_{\mathrm{DC}i}, U_{\mathrm{DC}})$ to $F(P_{\mathrm{DC}i}^{'}, U_{\mathrm{DC}}^{'})$, as shown in **Figure 3**.

It can be seen from **Eq. 6** that the DC side voltage change variable of the VSC is

$$\Delta U_{\mathrm{DC}} = U_{\mathrm{DC}}^{'} - U_{\mathrm{DC}} = -\frac{P_{\mathrm{DC}i}^{'} - P_{\mathrm{DC}i}}{K_{\mathrm{droop}i}} = -\frac{\Delta P_{\mathrm{DC}i}}{K_{\mathrm{droop}i}}. \tag{7}$$

After ignoring the DC line resistance, the whole DC network can be regarded as an equipotential point. Therefore, the total increment of the VSC output power can be expressed as

$$\Delta P_{\mathrm{DC}} = \sum_{i=1}^{n} \Delta P_{\mathrm{DC}i} = -\Delta U_{\mathrm{DC}} \sum_{i=1}^{n} K_{\mathrm{droop}i} = \frac{\Delta P_{\mathrm{DC}j}}{K_{\mathrm{droop}j}} \sum_{i=1}^{n} K_{\mathrm{droop}i}, \tag{8}$$

where $K_{\mathrm{droop}j}$ is the droop coefficient of VSC$_j$.

Therefore, the output power increment $\Delta P_{\mathrm{DC}j}$ of VSC$_j$ is

$$\Delta P_{\mathrm{DC}j} = K_{\mathrm{droop}j} \frac{\Delta P_{\mathrm{DC}}}{\sum_{i=1}^{n} K_{\mathrm{droop}i}}. \tag{9}$$

It can be seen from **Eq. 9** that when the total unbalanced power ΔP_{DC} of the VSC-MTDC power distribution system is constant, the unbalanced power borne by each VSC is proportional to its droop coefficient.

For conventional droop control, the droop coefficient is set proportional to the capacity of the VSC and remains constant during the operation process. When a considerable large disturbance occurs in the system, the VSC with a smaller capacity can be overloaded (corresponding to the operation point $G(P_{i\mathrm{MAX}}, U_{\mathrm{DC}}^{''})$), while the VSC with a larger capacity still has a certain margin (corresponding to the operation point $G(P_{\mathrm{DC}j}^{''}, U_{\mathrm{DC}}^{''})$). If the power balance of each VSC is broken, make the droop coefficient set by each VSC disobey the strict proportional relation, and the droop coefficient of the VSC with a larger power margin is improved to assume more unbalanced power during the disturbance, the droop coefficient of the VSC with a smaller power margin is appropriately reduced to assume less unbalanced power during the disturbance, then the unbalanced power is reasonably transferred, and the response capability of the VSC adopting the droop control strategy to the DC power flow disturbance can be indirectly promoted. In addition, the traditional droop control can realize the real-time regulation of the DC side voltage of the VSC with power changing according to the droop characteristic. However, when there is no power fluctuation in the transmittance VSC, the acceptance VSC still performs the differential adjustment according to droop characteristics, and it is impossible to reasonably realize the power distribution and to cause the DC voltage to fluctuate. Consequently, this article proposes an adaptive droop control strategy of the VSC-MTDC power distribution network taking

into account the power–voltage deviation, which can improve the response capability to DC power flow disturbance and at the same time can realize the isochronous control of DC voltage.

An improved droop coefficient that takes into account the power margin of the VSC is defined as

$$K_{\text{droop}i}^{'}=\begin{cases}\dfrac{\mu\left(P_{i\text{MAX}}-|P_{\text{DC}i}|\right)}{P_{i\text{MAX}}}K_{\text{droop}i}, U_{\text{DC}i}\geq U_{\text{DC}i}^{ref}\\ \dfrac{\mu|P_{\text{DC}i}|}{P_{i\text{MAX}}}K_{\text{droop}i}, U_{\text{DC}i}<U_{\text{DC}i}^{ref}.\end{cases}\tag{10}$$

and satisfies that

$$\sum_{i=1}^{n}K_{\text{droop}i}^{'}=\sum_{i=1}^{n}K_{\text{droop}i}.\tag{11}$$

In **Eq. 10**, μ is as a constant, responsible for the proper scaling of $K_{\text{droop}i}^{'}$, and its values are generally ranging within the region (Liao et al., 2018; Li et al., 2021a) according to the actual operation state of the power network (Tao et al., 2018); in this article, μ is equal to 3; $K_{\text{droop}i}$ is the traditional droop control coefficient, and that

$$K_{\text{droop}i}=\frac{P_{i\text{MAX}}-P_{\text{DC}i}}{U_{\text{DC}}-U_{\text{DCL}}}.\tag{12}$$

After the optimization by **Eq. 10**, the droop coefficient of VSC$_i$ decreases and the droop coefficient of VSC$_j$ becomes larger. Under the condition of a constant reference voltage, the optimal operating state points of the two VSCs, respectively, correspond to the points of $M\left(P_{\text{DC}i}^{'ref},U_{\text{DC}}^{ref}\right)$ and $M^{'}\left(P_{\text{DC}j}^{'ref},U_{\text{DC}}^{ref}\right)$ in **Figure 3**, thus realizing the optimal allocation of unbalanced power.

However, when the droop coefficient of the VSC with a smaller power margin is too low, the slight power fluctuation will lead to a large deviation between $U_{\text{DC}i}$ and U_{DC}^{ref}, which greatly increases the difficulty of DC voltage control and is not conducive to system stability, so a reasonable limit should be imposed on $K_{\text{droop}i}^{'}$, thus setting that (Tao et al., 2018)

$$\begin{cases}K_{\text{droop}i\text{MAX}}^{'}=\mu K_{\text{droop}i},\\ K_{\text{droop}i\text{MIN}}^{'}=\dfrac{1}{3}K_{\text{droop}i},\\ K_{\text{droop}i\text{MIN}}^{'}\leq K_{\text{droop}i}^{'}\leq K_{\text{droop}i\text{MAX}}^{'}.\end{cases}\tag{13}$$

At the same time, in order to ensure the continuity of DC voltage in the process of droop controlling, a power deviation factor ζ_i $(0<\zeta_i<1)$ is introduced, and the "dead zone" is properly set in the droop characteristic curve, perpendicular to the voltage shaft in **Figure 3**, so

$$\Delta P_{\text{DC}}=\sum_{i=1}^{n}\Delta P_{\text{DC}i}^{''}=\sum_{i=1}^{n}\zeta_i\Delta P_{\text{DC}}.\tag{14}$$

For VSC$_i$ in **Figure 3**, the adjusted droop characteristic curve is equivalent to translating the original curve to the right for $\zeta_i\Delta P_{\text{DC}}$, its assumed power increment becomes $\Delta P_{\text{DC}i}^{''}$, the steady operating state point is $M\left(P_{\text{DC}i}^{'ref},U_{\text{DC}}^{ref}\right)$, which only considers that the power margin is shifted to the point $K\left(P_{\text{DC}i}^{''ref},U_{\text{DC}}^{ref}\right)$ without the reference voltage changing; at present, the corresponding reference power increase is $\zeta_i\Delta P_{\text{DC}}$ up to $P_{\text{DC}i}^{''ref}$, where the inflection point C is a voltage deviation control enabling node, and

$$P_{\text{DC}i}^{''ref}=-\frac{U_{\text{DC}}^{ref}-U_{\text{DCL}}}{K_{\text{droop}i}^{'}\big|_{U_{\text{DC}i}=U_{\text{DC}}^{ref}}}+P_{\text{DC}i}^{ref}+\zeta_i\Delta P_{\text{DC}},\tag{15}$$

where

$$K_{\text{droop}i}^{'}\big|_{U_{\text{DC}i}=U_{\text{DC}}^{ref}}=\frac{\mu\left(P_{i\text{MAX}}-P_{\text{DC}i}^{ref}\right)^2}{P_{i\text{MAX}}\left(U_{\text{DC}}^{ref}-U_{\text{DCL}}\right)}.$$

After the additional DC voltage deviation control, the droop coefficient of the controller (corresponding to the slope of the characteristic curve of droop control) holds in line, and the power reference value of the optimal operating state point increases, which expands the operation range of VSC to a certain extent, alleviates the power margin decrease of the VSC in the case of only adopting power margin control, and can obviously enhance system voltage stability. The adaptive droop control block diagram of the double closed loop based on the PI link is shown in **Figure 4**.Here, $U_{\text{DC}r}$ is the DC voltage modulation value; $K_{\text{P}i}$, $K_{\text{I}i}$, $K_{\text{P}i}^{'}$, and $K_{\text{I}i}^{'}$ are the PI controller coefficients; $K_{\text{GU}i}$ and $K_{\text{GD}i}$ are the control identification bits; and the corresponding control mode is enabled when the value is 1.

RESPONSE CHARACTERISTIC ANALYSIS OF ADAPTIVE DROOP CONTROL

To clarify the relationship between the output DC voltage and power of each VSC under the adaptive droop control in the VSC-MTDC distribution network, it is necessary to analyze the U–P response characteristics of VSCs. **Figure 4** shows that when the control identification bit coefficient $K_{\text{GU}i}$ and $K_{\text{GD}i}$ are both 1, the output power–voltage relationship of the VSC is

$$\begin{aligned}&P_{\text{DC}i}^{'ref}-P_{\text{DC}i}+\left(K_{\text{droop}i}^{'}+1\right)\left(U_{\text{DC}}^{ref}-U_{\text{DC}i}\right)\\&=\frac{1}{K_{\text{P}i}+\frac{K_{\text{I}i}}{s}}\left(\frac{1}{K_{\text{P}i}^{'}+\frac{K_{\text{I}i}^{'}}{s}}U_{\text{DC}r}+I_{\text{DC}i}\right).\end{aligned}\tag{16}$$

Because the response speed of outer loop voltage control and inner loop current control is much higher than that of droop control, the DC voltage stability of the system is less affected by the parameters of the PI controller and more significantly affected by the droop coefficient (Liu et al., 2019). Therefore, assuming that the closed loop transfer function of DC voltage is 1, there is

$$U_{\text{DC}r}=U_{\text{DC}i}.\tag{17}$$

With the introduction of unit step response into the steady operation of the VSC-MTDC distribution network, that is, the system power demand suddenly increases to 1 kW, the response relationship of the VSC-MTDC distribution system under different droop coefficients $K_{\text{droop}i}^{'}$ is obtained by **Eqs. 16**, **17** and shown in **Figure 5**, and the response relationship scheme is

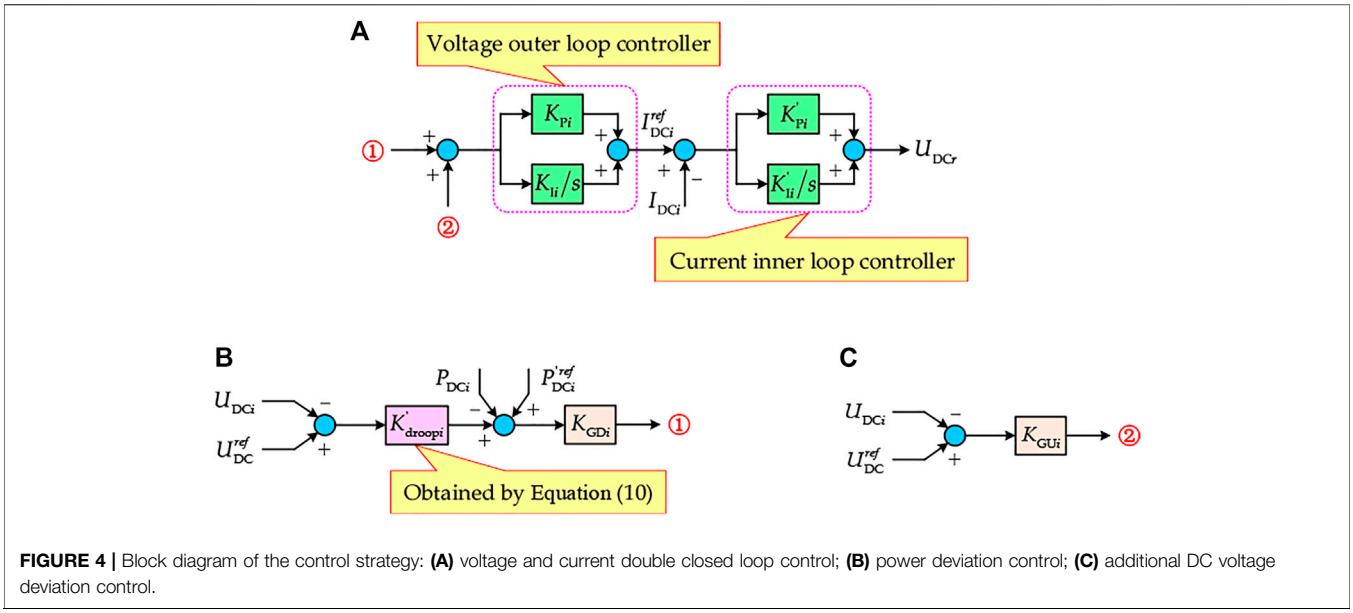

FIGURE 4 | Block diagram of the control strategy: **(A)** voltage and current double closed loop control; **(B)** power deviation control; **(C)** additional DC voltage deviation control.

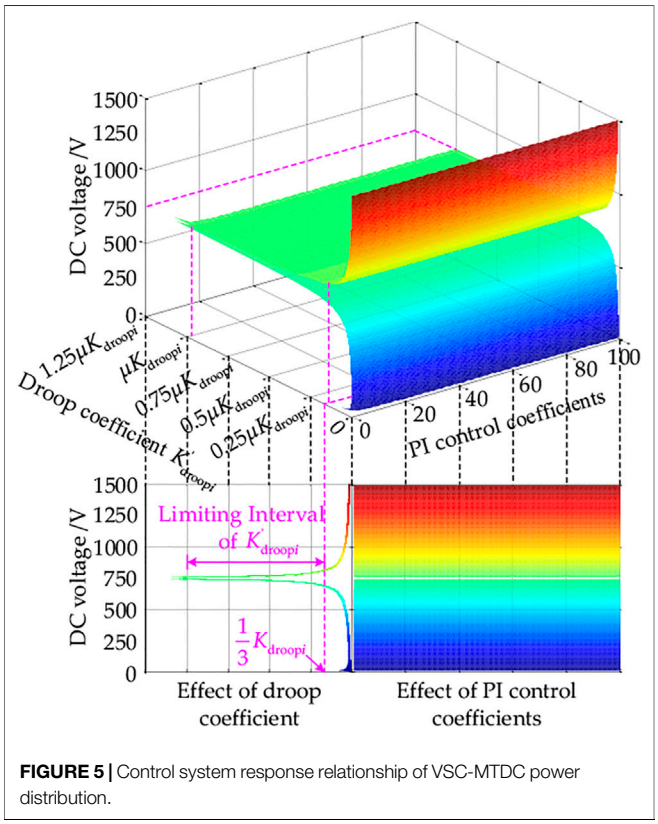

FIGURE 5 | Control system response relationship of VSC-MTDC power distribution.

segmented with the DC voltage reference value (U_{DC}^{ref} = 750V) as the critical point. In the limiting range of the droop coefficient $K_{\mathrm{droop}i}^{'}$, the DC side voltage of the VSC can basically change in the range of ±5% of U_{DC}^{ref}, and for any VSC, the larger the droop coefficient is, the stronger the stability of the system is.

TABLE 1 | Parameters of response characteristic analysis.

Parameters	Data
DC voltage reference value U_{DC}	750 V
Output power reference value $P_{\mathrm{DC}i}^{'ref}$	80kW
Outer loop control coefficient: $K_{\mathrm{P}i}$ and $K_{\mathrm{I}i}$	0.045 and 138
Inner loop control coefficient: $K_{\mathrm{P}i}^{'}$ and $K_{\mathrm{I}i}^{'}$	0.015 and 105

The set VSC-MTDC distribution network control parameters for response characteristic analysis are shown in **Table 1**. Here, the outer and inner loop control coefficients are calculated according to the method of Wang et al., 2018, U_{DC} and $P_{\mathrm{DC}i}$ are chosen according to their reference values, the maximum operating power of the VSC is selected by 70% of the reference capacity, and the lower limit of DC voltage of the VSC is less than 30% of the reference voltage; therefore,

$$K_{\mathrm{droop}i} = \frac{80000 \times (1 + 70\%) - 80000}{750 - 750 \times (1 - 30\%)} = 106.67, \tag{18}$$

$$\begin{cases} K_{\mathrm{droop}i\mathrm{MAX}}^{'} = \mu K_{\mathrm{droop}i} = 3 \times 106.67 = 320 \\ K_{\mathrm{droop}i\mathrm{MIN}}^{'} = \frac{1}{3} K_{\mathrm{droop}i} = \frac{1}{3} \times 106.67 = 35.56. \end{cases} \tag{19}$$

Therefore, the limiting interval of the droop coefficient is determined to be [35, 320].

When the power fluctuation occurs, to ignore the quadratic disturbance term and to linearize **Eq. 16**, the small-signal closed-loop transfer function $T(s)$ of adaptive droop control can be expressed as follows

$$T(s) = \frac{\Delta U_{\mathrm{DC}i}(s)}{\Delta P_{\mathrm{DC}i}(s)} = \frac{b_2 s^2 + b_1 s}{a_2 s^2 + a_1 s + a_0}, \tag{20}$$

where

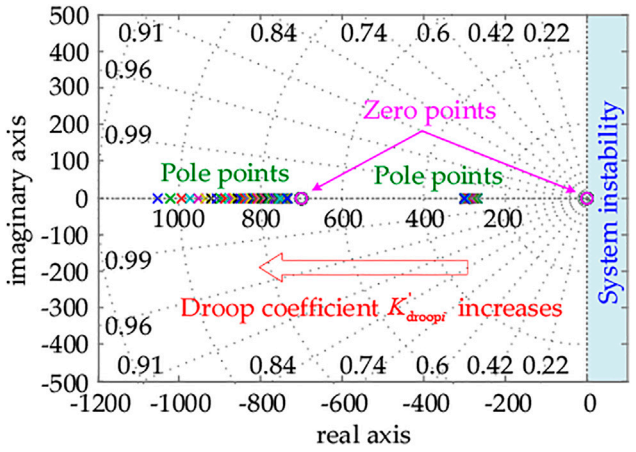

FIGURE 6 | Zero-pole distribution map of $T(s)$ under droop coefficient variation.

$$\begin{cases} a_0 = K_{Ii}K'_{Ii}\left(U^{ref}_{DC}K'_{droopi} + U^{ref}_{DC} + P'^{ref}_{DCi}\right), \\ a_1 = \left(K_{Ii}K'_{Pi} + K_{Pi}K'_{Ii}\right)\left(U^{ref}_{DC}K'_{droopi} + U^{ref}_{DC} + P'^{ref}_{DCi}\right), \\ a_2 = K_{Pi}K'_{Pi}\left(U^{ref}_{DC}K'_{droopi} + U^{ref}_{DC} + P'^{ref}_{DCi}\right) - U^{ref}_{DC}, \\ b_1 = K'_{Ii}, \\ b_2 = K'_{Pi}. \end{cases}$$

Figure 6 shows the zero-pole distribution map of the transfer function $T(s)$. In the limiting range of the droop coefficient K'_{droopi}, all poles are in the left half of the complex plane and on the real axis, so the VSC-MTDC power distribution network is always stable, and the system stability is independent of the inner and outer loop control coefficients and the power and voltage reference values. Only when K'_{droopi} reduces to a value far below its minimum value, the poles may appear to the right of the imaginary axis, at which point the system will be unstable. As shown in **Figure 7**, when the inner or outer loop control coefficients change, the pole will shift on the negative half axis of the real axis, and the larger the pole value is, the farther away the pole is from the virtual axis and the faster the response speed of the system is.

SIMULATION VERIFICATION AND ANALYSIS

Parameters of the Simulation Model

In order to verify the effectiveness and control effect of adaptive droop control proposed in this article, the electromagnetic transient model of the five-terminal VSC-MTDC distribution network is established on the MATLAB/Simulink software platform, in which the other four VSCs all perform adaptive droop control considering the power–voltage deviation from the constant DC voltage control adopted in VSC_5. In this section, the simulation experiments are carried out for three operating conditions, including the equivalent load fluctuation of the DC network, VSC_3 with droop control exiting operation, and VSC_5 with fixed DC voltage control exiting operation, and the simulation results are compared and analyzed in detail with traditional droop control. The main parameters of the simulation model are shown in **Table 2**.

Analysis of Simulation Results

Operating Condition 1: Equivalent Load Fluctuation in the DC Network

When t = 1.4s was set, the equivalent load of the DC network increased from 500kW to 545kW, the resulting power shortage led to the decrease of DC bus voltage, and then each VSC increased power output to maintain the stability of DC bus voltage.

Under the traditional droop control, the power shortage of the system should be allocated strictly according to the capacity of the VSC. As shown in **Figure 8A**, each VSC, that is, VSC_1, VSC_2, VSC_3, and VSC_4, which adopted the droop control strategy, respectively, increased the power output by 6.67 kwW, 13.33 kwW, 10.62 kW, and 9.38 kW, and the homologous load rates were 82.72, 60.49, 56.10, and 49.67% apart; VSC_1 operated with heavy load. The time for the system to recover stability is more than 1.6s, and there was a DC bus voltage deviation of 8.52 V compared with 1.4s ago, and the voltage deviation rate was 1.14%. During the period, the peak voltage fluctuation of the DC bus reached 54.55V, accounting for 7.27% of the rated voltage. As shown in **Figure 8B**, under the adaptive droop control, the unbalanced power borne by each VSC with droop control broke the fixed proportional constraint and, respectively,

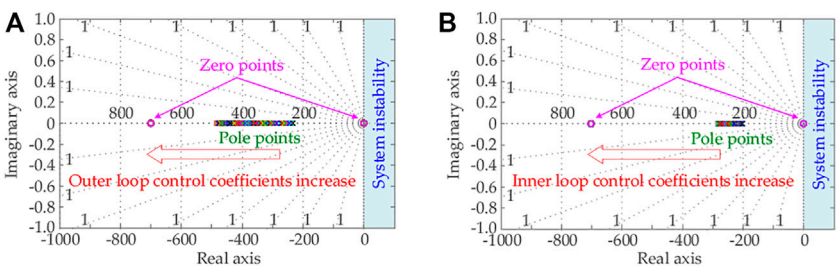

FIGURE 7 | Zero-pole distribution map of $T(s)$: **(A)** zero-pole distribution map of outer loop control coefficient variation; **(B)** zero-pole distribution map of inner loop control coefficient variation.

TABLE 2 | Parameters of response characteristic analysis.

Parameters	Data
Rated primary voltage of the AC system	10.5kV
Rated secondary voltage of the AC system	0.4kV
Rated capacity of the AC system	300 kVA
Ratio of equivalent reactance to resistance in the AC system	5
Rated voltage of the DC network	750 V
Rated capacity of VSCs with droop control	135, 270, 215, and 190 kVA
DC side capacitance of VSC	5000 μF
DC side flat wave reactance of VSC	0.2mH
Initial equivalent load of the DC network	500kW

increased the power output of 3.27 kW, 21.16 kW, 10.63 kW, and 8.65 kW, and the corresponding load rates were 80.20, 63.39, 56.11, and 49.29%, respectively, and the load rate of VSC_1 obviously reduced. The system restored stability after 0.14s, during which the peak voltage fluctuation of the DC bus reached 14.10V, accounting for 1.88% of the rated voltage.

Operating Condition 2: VSC With Droop Control Exiting Operation

When t = 1.4s was set, with VSC_3 with droop control exited operation, its power output reduced to 0, the resulting power shortage of 110 kW substantially led to the decrease of DC bus voltage, and then other VSCs increased the power output to maintain the stability of DC bus voltage.

As shown in **Figure 9A**, under the traditional droop control, the power output increments of the VSC_1, VSC_2, and VSC_4 adopted droop control strategies were 22.96kW, 45.95, and 32.34kW apart, and the homologous load rates were 94.79, 72.57, and 61.76%, respectively, VSC_1 is close to the full load, and there is a great operational risk. There was a DC bus voltage deviation of about 9.55 V between 1.4s before and after, and the voltage deviation rate was 1.27%. During the period, the peak voltage fluctuation of the DC bus reached 95.45V, accounting for 12.73% of the rated voltage. As shown in **Figure 9B**, under the adaptive droop control, the power output increments of the VSC_1, VSC_2, and VSC_4 adopted droop control strategies were 12.49kW, 56.16 and 38.55kW, respectively, and the corresponding load rates were 87.03, 76.36, and 65.03%, respectively, and the load rate of VSC_1 obviously reduced. The system restored stability after 0.19 s, during which the peak voltage fluctuation of the DC bus reached 30.60V, accounting for 4.08% of the rated voltage.

Operating Condition 3: VSC With Fixed DC Voltage Control Exiting Operation

When t = 1.4s was set, with VSC_5 with fixed DC voltage control exited operation, its power output reduced to 0, the resulting power shortage of 50 kW substantially led to the decrease of DC bus voltage, and then other four VSCs increased the power output to maintain the stability of DC bus voltage.

As shown in **Figure 10A**, under the traditional droop control, the power output increments of the VSC_1, VSC_2, VSC_3, and VSC_4 adopted droop control strategies were 8.33kW, 16.67kW, 13.27kW, and 11.73kW apart and the homologous load rates were 83.95, 61.73, 57.33, and 50.91%, respectively; VSC_1 operated with heavy load. The time for the system to recover stability is more than 1.6 s due to the loss of DC voltage support; after stabilization, there will still be a certain deviation from that before 1.4s. During the period, the DC bus voltage fluctuated violently with a peak value of 156.82V, accounting for 20.91% of the rated

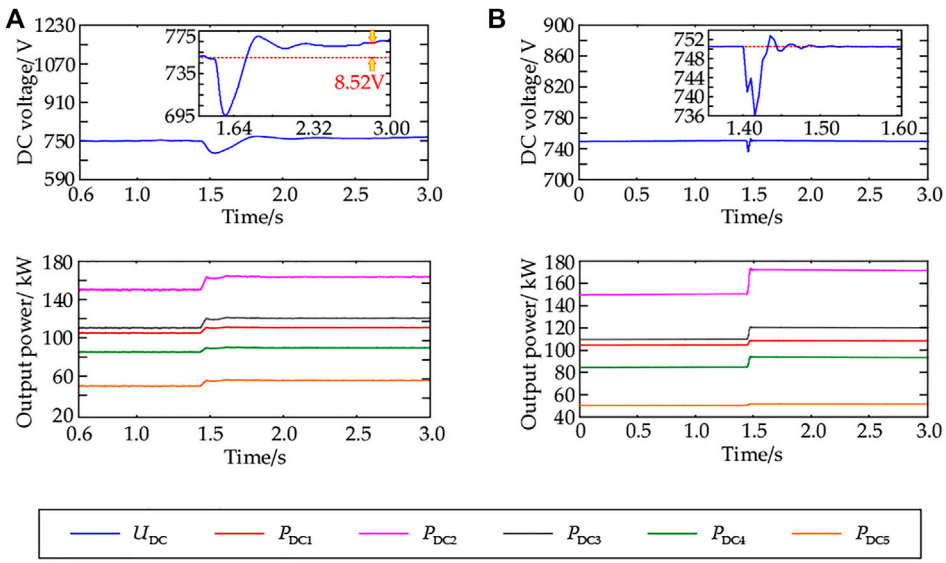

FIGURE 8 | Simulation waveforms of equivalent load fluctuation in the DC network: **(A)** simulation waves based on traditional droop control; **(B)** simulation waves based on adaptive droop control.

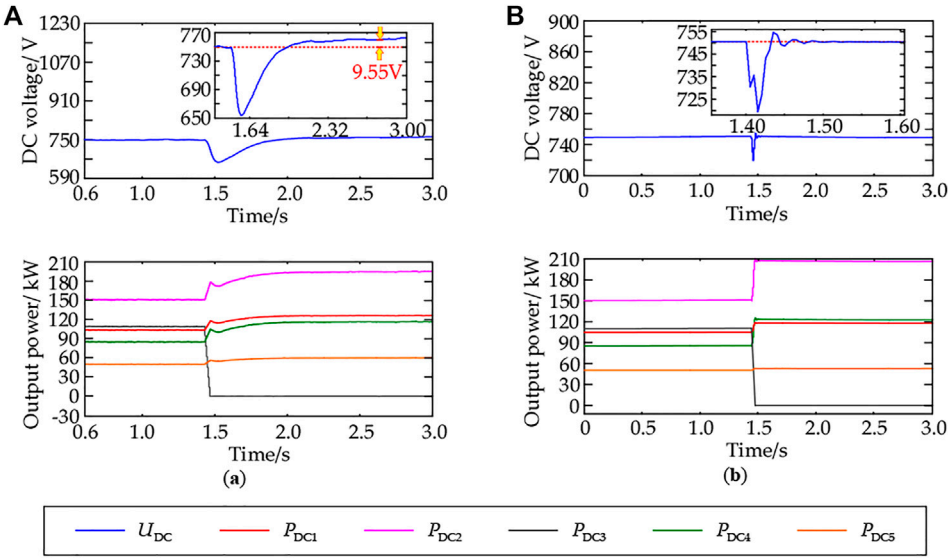

FIGURE 9 | Simulation waveforms of the VSC with droop control exiting operation: **(A)** simulation waves based on traditional droop control; **(B)** simulation waves based on adaptive droop control.

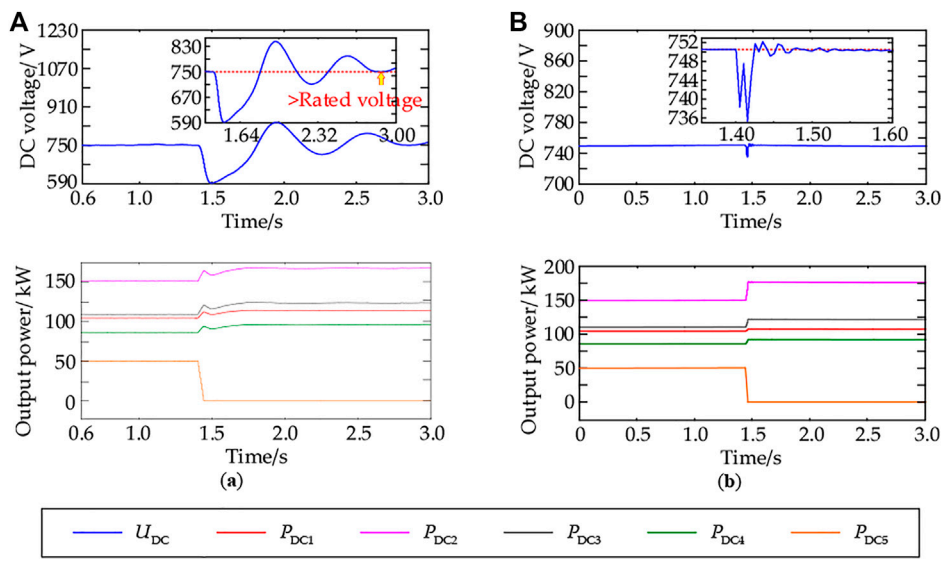

FIGURE 10 | Simulation waveforms of the VSC with fixed DC voltage control exiting operation: **(A)** simulation waves based on traditional droop control; **(B)** simulation waves based on adaptive droop control.

voltage. As shown in **Figure 10B**, under the adaptive droop control, the power output increments of VSC_1, VSC_2, VSC_3, and VSC_4 were 4.76kW, 25.32kW, 11.98kW, and 7.94kW, respectively, and the corresponding load rates were 81.30, 64.93, 56.73, and 48.92%; the load rate of the VSC_1 obviously reduced. Because of the addition of the DC voltage deviation control in the droop characteristic, the ability to restore the system's stable state obviously enhanced, the time consumption is about 0.39s, and the peak value of the DC bus voltage fluctuation during the period is 14.70V, accounting for 1.96% of the rated voltage.

CONCLUSION

In this article, the adaptive drooping characteristic optimization method considering power–voltage deviation is applied to the VSC-MTDC distribution network. The

proposed control strategy is modeled and simulated based on MATLAB/Simulink under different system operation conditions and compared with the traditional droop control. The conclusions are as follows:

1) Based on the analysis of the response characteristics of the VSC-MTDC distribution network under the proposed control strategy, the stability of the control system in the limiting range of droop coefficient is verified.
2) The active power optimal allocation is realized between each VSC, and when the system is disturbed, the average load rate of the VSC with a small capacity reduces by about 6.59%, the overload risk debases, and the response ability of the VSC to DC power flow disturbance obviously improves.
3) The system DC voltage deviation before and after disturbance caused by the differential control characteristic of the conventional droop control is basically eliminated, and the isochronous control to DC voltage is achieved. At the same time, the average recovery time of the system is shortened more than 78.74% and the average transient voltage peak value during the period reduces about 76.71%, which greatly improves the reliability and power quality of power supply to users.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/Supplementary Material; further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

YL and ZW contributed to conceptualization. YZ and LC contributed to methodology. YZ contributed to software. ZW contributed to validation. JZ and HL contributed to formal analysis. YL and QK contributed to resources. YL and YZ contributed to data curation. YL and ZW contributed to project administration.

FUNDING

This study received funding from Science and Technology Project of State Grid Corporation (No. SGTYHT/21-JS-225). The funder had the following involvement with the study: conceptualization, formal analysis, resources, data curation and project administration.

REFERENCES

Beerten, J., and Belmans, R. (2013). Analysis of Power Sharing and Voltage Deviations in Droop-Controlled DC Grids. *IEEE Trans. Power Syst.* 28, 4588–4597. doi:10.1109/tpwrs.2013.2272494

Chen, J. K., Dong, F. F., Wang, Z. H., Li, G. Q., and Zhang, C. (2018). Research on Improved Droop Control Method of Multi–Terminal MMC–HVDC System Suitable for Power Fluctuation. *J. Power Syst. Tech.* 42, 3708–3717.

Gao, Y., Ai, Q., Yousif, M., and Wang, X. (2019). Source-load-storage Consistency Collaborative Optimization Control of Flexible DC Distribution Network Considering Multi-Energy Complementarity. *Int. J. Electr. Power Energ. Syst.* 107, 273–281. doi:10.1016/j.ijepes.2018.11.033

Han, M. X., Xu, D., and Wan, L. (2016). Consensus Algorithm Based Decentralized Autonomous Control of Hybrid Multi–Terminal Direct Current System. *J. Automation Electric Power Syst.* 40, 130–136.

Jaynendra, K., Anshul, A., and Vineeta, A. (2019). A Review on Overall Control of DC Microgrids. *J. J. Energ. Storage* 21, 113–138.

Li, B., Liu, T., and Zhang, Y. (2017). Unified Adaptive Droop Control Design Based on Dynamic Reactive Power Limiter in VSC-MTDC. *Electric Power Syst. Res.* 148, 18–26. doi:10.1016/j.epsr.2017.03.010

Li, H. L., Jiang, J. G., Zhou, Z. Z., and Zhang, D. (2019). Scheme and Control Method of MMC Based Medium–Voltage and High–Power DC Distribution System. *J. Automation Electric Power Syst.* 43, 83–88.

Li, W. G., and Lao, X. T. (2017). Optimized Distribution for Active Power in Parallel AC/DC Transmission Systems. *J. J. Northeast Electric Power Univ.* 37, 24–30.

Li, Y., Han, M., Yang, Z., and Li, G. (2021). Coordinating Flexible Demand Response and Renewable Uncertainties for Scheduling of Community Integrated Energy Systems with an Electric Vehicle Charging Station: A Bi-level Approach. *IEEE Trans. Sustain. Energ.* 12, 2321–2331. doi:10.1109/tste.2021.3090463

Li, Y., Li, K., Yang, Z., Yu, Y., Xu, R., and Yang, M. (2022). Stochastic Optimal Scheduling of Demand Response-Enabled Microgrids with Renewable Generations: An Analytical-Heuristic Approach. *J. Clean. Prod.* 330, 129840. doi:10.1016/j.jclepro.2021.129840

Li, Y., Wang, C., Li, G., and Chen, C. (2021). Optimal Scheduling of Integrated Demand Response-Enabled Integrated Energy Systems with Uncertain Renewable Generations: A Stackelberg Game Approach. *Energ. Convers. Manag.* 235, 113996. doi:10.1016/j.enconman.2021.113996

Liao, J. Q., Zhou, N. C., Wang, Q. G., Li, C. Y., and Yang, J. (2018). Definition and Correlation Analysis of Power Quality index of DC Distribution Network. *J. Proc. CSEE* 38, 6847–6860.

Liu, Z. W., Miao, S. H., Fan, Z. H., Chao, K. Y., and Kang, Y. L. (2019). Accurate Power Allocation and Zero Steady–State Error Voltage Control of the Islanding DC Microgird Based on Adaptive Droop Characteristics. *J. Trans. China Electrotechnical Soc.* 34, 795–806.

Ma, D., Cao, X., Sun, C., Wang, R., Sun, Q., Xie, X., et al. (2021). "Dual-predictive Control with Adaptive Error Correction Strategy for AC Microgrids," in *IEEE Trans. Power Delivery* (IEEE), 1. doi:10.1109/TPWRD.2021.3101198

Pedram, G., and Mohsen, N. (2018). A Distributed Control Strategy Based on Droop Control and Low–Bandwidth Communication in DC Microgrids with Increased Accuracy of Load Sharing. *J. Sust. Cities Soc.* 40, 155–164.

Qusay, S., and Xie, J. (2018). Decentralized Power Control Management with Series Transformer Less H–Bridge Inverter in Low–Voltage Smart Microgrid Based P–V Droop Control. *J. Electr. Power Energ. Syst.* 99, 500–515.

Rouzbehi, K., Miranian, A., Luna, A., and Rodriguez, P. (2014). DC Voltage Control and Power Sharing in Multiterminal DC Grids Based on Optimal DC Power Flow and Voltage-Droop Strategy. *IEEE J. Emerg. Sel. Top. Power Electron.* 2, 1171–1180. doi:10.1109/jestpe.2014.2338738

Tao, Y., Liu, T. Q., Li, B. H., Miao, D., Dong, Y. Q., and Lu, Z. X. (2018). Hierarchical Coordinated Adaptive Droop Control in Flexible HVDC Grid. *J. Automation Electric Power Syst.* 42, 70–79.

Wang, R., Sun, Q., Hu, W., Li, Y., Ma, D., and Wang, P. (2021). SoC-based Droop Coefficients Stability Region Analysis of the Battery for Stand-Alone Supply Systems with Constant Power Loads. *IEEE Trans. Power Electron.* 36, 7866–7879. doi:10.1109/tpel.2021.3049241

Wang, R., Sun, Q., Ma, D., and Liu, Z. (2019). The Small-Signal Stability Analysis of the Droop-Controlled Converter in Electromagnetic Timescale. *IEEE Trans. Sustain. Energ.* 10, 1459–1469. doi:10.1109/tste.2019.2894633

Wang, R., Sun, Q., Tu, P., Xiao, J., Gui, Y., and Wang, P. (2021). Reduced-order Aggregate Model for Large-Scale Converters with Inhomogeneous Initial

Conditions in DC Microgrids. *IEEE Trans. Energ. Convers.* 36, 2473–2484. doi:10.1109/tec.2021.3050434

Wang, W., and Barnes, M. (2014). Power Flow Algorithms for Multi-Terminal VSC-HVDC with Droop Control. *IEEE Trans. Power Syst.* 29, 1721–1730. doi:10.1109/tpwrs.2013.2294198

Wang, Y., He, J., Zhao, Y., Liu, G., Sun, J., Li, H., et al. (2020). Equal Loading Rate Based Master-Slave Voltage Control for VSC Based DC Distribution Systems. *IEEE Trans. Power Deliv.* 35, 2252–2259. doi:10.1109/tpwrd.2020.2964706

Wang, Y., Li, B., Zhou, Z., Chen, Z., Wen, W., Li, X., et al. (2020). DC Voltage Deviation-dependent Voltage Droop Control Method for VSC-MTDC Systems under Large Disturbances. *IET Renew. Power Generation* 14, 891–896. doi:10.1049/iet-rpg.2019.0394

Wang, Y., Wen, W., Wang, C., Liu, H., Zhan, X., and Xiao, X. (2019). Adaptive Voltage Droop Method of Multiterminal VSC-HVDC Systems for DC Voltage Deviation and Power Sharing. *J. IEEE Trans. Power Deliv.* 34, 169–176.

Wang, Z. X., Li, Y. Z., and Li, G. Q. (2018). Parameters Optimization of Double Closed–Loop for LCL–type Inverter Based on Genetic Algorithm. *J. Power Syst. Prot. Control.* 46, 1–7.

Xu, F., Guo, Q., Sun, H., Zhang, B., and Jia, L. (2019). A Two-Level Hierarchical Discrete-Device Control Method for Power Networks with Integrated Wind Farms. *J. Mod. Power Syst. Clean. Energ.* 7, 88–98. doi:10.1007/s40565-018-0417-1

Yang, N., Paire, D., Gao, F., Miraoui, A., and Liu, W. (2015). Compensation of Droop Control Using Common Load Condition in DC Microgrids to Improve Voltage Regulation and Load Sharing. *Int. J. Electr. Power Energ. Syst.* 64, 752–760. doi:10.1016/j.ijepes.2014.07.079

Zhao, X. S., Peng, K., Zhang, X. H., Xu, B. Y., Chen, Y., and Zhao, Y. H. (2019). Analysis on Dynamic Performance of Droop Control for Multi–Terminal VSC Based DC Distribution System. *J. Automation Electric Power Syst.* 43, 89–96.

Conflict of Interest: Authors YL, JZ, HL and QK are employed by Smart Distribution Network Center, State Grid Jibei Electric Power Co., Ltd & State Grid Corporation.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Advantages of publishing in Frontiers

OPEN ACCESS
Articles are free to read for greatest visibility and readership

FAST PUBLICATION
Around 90 days from submission to decision

HIGH QUALITY PEER-REVIEW
Rigorous, collaborative, and constructive peer-review

TRANSPARENT PEER-REVIEW
Editors and reviewers acknowledged by name on published articles

Frontiers
Avenue du Tribunal-Fédéral 34
1005 Lausanne | Switzerland

Visit us: www.frontiersin.org
Contact us: frontiersin.org/about/contact

REPRODUCIBILITY OF RESEARCH
Support open data and methods to enhance research reproducibility

DIGITAL PUBLISHING
Articles designed for optimal readership across devices

FOLLOW US
@frontiersin

IMPACT METRICS
Advanced article metrics track visibility across digital media

EXTENSIVE PROMOTION
Marketing and promotion of impactful research

LOOP RESEARCH NETWORK
Our network increases your article's readership