frontiers | Frontiers in Energy Research

TYPE Original Research
PUBLISHED 08 October 2024
DOI 10.3389/fenrg.2024.1448046

OPEN ACCESS

EDITED BY
C. H. Rami Reddy,
Chonnam National University, Republic of Korea

REVIEWED BY
Manne Bharathi,
Acharya Nagarjuna University, India
Linfei Yin,
Guangxi University, China
Lefeng Cheng,
Guangzhou University, China

*CORRESPONDENCE
Yongzhi Li,
tuytliyongzhi@163.com

RECEIVED 12 June 2024
ACCEPTED 30 August 2024
PUBLISHED 08 October 2024

CITATION
Xiao J, Zhao W, Li W, Zhao Y, Li Y, Ma X and Liu Y (2024) Active power balance control of wind-photovoltaic-storage power system based on transfer learning double deep Q-network approach.
Front. Energy Res. 12:1448046.
doi: 10.3389/fenrg.2024.1448046

Active power balance control of wind-photovoltaic-storage power system based on transfer learning double deep Q-network approach

Jun Xiao, Wen Zhao, Wei Li, Yankai Zhao, Yongzhi Li*, Xudong Ma and Yuchao Liu

State Grid Shanxi Power Company Lvliang Power Supply Company, Lvliang, China

Introduction: This study addresses the challenge of active power (AP) balance control in wind-photovoltaic-storage (WPS) power systems, particularly in regions with a high proportion of renewable energy (RE) units. The goal is to effectively manage the AP balance to reduce the output of thermal power generators, thereby improving the overall efficiency and sustainability of WPS systems.

Methods: To achieve this objective, we propose the transfer learning double deep Q-network (TLDDQN) method for controlling the energy storage device within WPS power systems. The TLDDQN method leverages the benefits of transfer learning to quickly adapt to new environments, thereby enhancing the training speed of the double deep Q-network (DDQN) algorithm. Additionally, we introduce an adaptive entropy mechanism integrated with the DDQN algorithm, which is further improved to enhance the training capability of agents.

Results: The proposed TLDDQN algorithm was applied to a regional WPS power system for experimental simulation of AP balance control. The results indicate that the TLDDQN algorithm trains agents more rapidly compared to the standard DDQN algorithm. Furthermore, the AP balance control method based on TLDDQN can more accurately manage the storage device, thereby reducing the output of thermal power generators more effectively than the particle swarm optimization-based method.

Discussion: Overall, the TLDDQN algorithm proposed in this study can provide some insights and theoretical references for research in related fields, especially those requiring decision making.

KEYWORDS

wind-photovoltaic-storage power system, renewable energy, active power balance control, double deep Q-Network, transfer learning

1 Introduction

Conventional power generation technologies produce large amounts of greenhouse gases (Russo et al., 2023). To reduce greenhouse gas emissions, various countries have formulated carbon reduction programs. Renewable energy (RE) power generation technology has been widely favored by countries for the advantages of environmental

protection and sustainability (Han et al., 2023). However, the stochastic and fluctuating characteristics of RE generation systems can threaten the reliability of power systems (Guerra et al., 2022). Energy storage (ES) devices can release power to relieve power tension or absorb power to avoid power waste (Dong et al., 2022). Consequently, the stability of the RE power generating system can be enhanced by the RE power plant built by leveraging the complementarity of RE.

When the proportion of RE units in power generation systems is small, the traditional active power (AP) regulation strategy of the RE power generation system can prioritize the consumption of power generated by RE units. The thermal power units cooperate with the RE units to regulate the AP balance of the RE power generation system (Grover et al., 2022). However, when the proportion of RE units in the RE power generation system is large, the RE units need to cooperate with the traditional thermal power units to control the AP balance of the power system (Ye et al., 2023). In this study, the AP balance control problem is considered for a high percentage of RE generation systems.

The AP balance control methods of RE generation systems mainly have two categories: swarm intelligence algorithms and reinforcement learning algorithms. The adaptability of the swarm intelligence algorithm-based AP balance control method is considerable. However, the swarm intelligence algorithm-based AP balance control method has the disadvantages of poor real-time performance and easily falling into local optimization (Moosavian et al., 2024). On the contrary, the AP balance control method based on reinforcement learning has the advantage of high real-time performance (Yin and Wu, 2022).

The swarm intelligence algorithm-based AP balance control method has the advantage of adaptability (Jiang et al., 2022). The AP balance control methods, which are based on hybrid swarm intelligence algorithms comprising Mexican axolotl optimization and the honey badger algorithm, have the potential to reduce carbon emissions, power costs, and peak power consumption in power systems (Revathi et al., 2024). An integrated load scheduling method for RE generation systems based on the Firefly algorithm can reduce the fuel cost of the generation system (Mehmood et al., 2023). Optimal AP scheduling methods for power systems based on hybrid particle swarm optimization and hippocampus optimization algorithms can reduce AP losses in power systems (Hasanien et al., 2024). However, the AP balance control method of wind-photovoltaic-storage (WPS) power system based on swarm intelligence algorithms has the shortcomings of low real-time performance and insufficient regulation accuracy.

The reinforcement learning-based AP balance control method is suitable for AP balance control of power systems in complex environments (Cheng and Yu, 2019). In addition, the AP balance control method for WPS power systems based on reinforcement learning has the advantage of high real-time performance. A decomposed predictive fractional-order proportional-integral-derivative control reinforcement learning algorithm can reduce frequency deviation and improve power quality in integrated energy systems (Yin and Zheng, 2024). The short-term optimal dispatch model framework of the water-wind-photovoltaic multi-energy power system constructed based on the deep Q-network (DQN) algorithm can improve the generation efficiency of multi-energy systems (Jiang et al., 2023). The control strategy of ES devices for energy systems based on improved deep deterministic policy gradient algorithms can integrate the frequency fluctuation of energy systems (Yakout et al., 2023). The energy system optimization control strategy based on the twin delayed deep deterministic policy gradient algorithm can flexibly adjust the components' operation and the ES device's charging strategy according to the output of RE sources and the electricity price (Zhang et al., 2022). The approach of employing electric vehicles as energy storage devices and regulating charging strategies with DQN algorithms is an effective solution to address the security of energy supply issues associated with the future power grid (Hao et al., 2023). A multi-agent game operation strategy consisting of energy retailers, suppliers, and users with integrated demand response is an effective way to alleviate the tension of multi-energy coupling and multi-agent difficulties (Li et al., 2023). N Population multi-strategy evolutionary game theory reveals the long-term equilibrium properties of the long-term bidding problem on the generation side of the power market and provides a theoretical reference to the complex dynamic interactive decision-making problems in related fields (Cheng et al., 2020). However, previous AP balance control methods based on reinforcement learning often need to be relearned when faced with new environments.

This study proposes the transfer learning double deep Q-network (TLDDQN)-based AP balance control method for controlling storage devices in WPS power systems. The proposed TLDDQN combines the advantage of transfer learning that can rapidly adapt to new environments and the advantage of the double deep Q-network (DDQN) algorithm that deals with complex environments. In addition, this study proposes a method to combine the adaptive entropy mechanism to the DDQN algorithm and improve the corresponding adaptive entropy mechanism. Therefore, the TLDDQN method can be effective in training TLDDQN agents and controlling the AP of the WPS power system. The characteristics of the AP balance control method for WPS power systems based on the proposed TLDDQN can be summarized as follows.

(1) This study combines the transfer learning approach and the DDQN to form the TLDDQN algorithm. The proposed TLDDQN algorithm combines the adaptive entropy mechanism to enhance the exploration ability during training and utilizes the transfer learning approach to transfer the generic parameters in the neural network (NN) of the TLDDQN algorithm.
(2) The active probabilistic balance control method for WPS power systems based on the proposed TLDDQN can be applied to control ES devices in WPS power systems.
(3) The active probabilistic balancing control method of the WPS power system based on the TLDDQN algorithm can balance the AP of the WPS power system.

2 Mathematical modeling of renewable energy generators

The devices of the WPS power system are mainly composed of wind power (WP) generation devices, photovoltaic power (PP)

generation devices, and ES devices (Abdelghany et al., 2024). This study analyzes the output characteristics of WP generation devices, PP generation devices, and ES devices to obtain the corresponding mathematical model.

2.1 Mathematical modeling of wind power generation devices

The WP generation devices convert the kinetic energy of the wind into electrical energy (Liu and Wang, 2022). The power generation efficiency of a WP generation device is related to the ambient wind speed (Jung and Schindler, 2023). The output of WP generation devices is expressed as follows (Equation 1).

$$P_{\mathrm{wt}} = \begin{cases} 0, & v < v_{\mathrm{ci}}^{'} \\ a'v^3 + b'v^2 + c'v + d', & v_{\mathrm{ci}}^{'} \le v \le v_{\mathrm{r}}^{'} \\ P_{\mathrm{r}}^{'}, & v_{r}^{'} < v < v_{\mathrm{co}}^{'} \\ 0, & v \le v_{\mathrm{co}}^{'} \end{cases} \tag{1}$$

where, P_{wt} is the output of WP generation devices; $P_{\mathrm{r}}^{'}$ is the rated power of WP generation devices; v is the actual wind speed; $v_{\mathrm{ci}}^{'}$ is the tangential wind speed of WP generation devices; $v_{\mathrm{r}}^{'}$ is the rated wind speed; $v_{\mathrm{co}}^{'}$ is the cut-out wind speed; a', b', c' and d' are the wind speed parameters of WP generation devices.

2.2 Mathematical modeling of photovoltaic power generation devices

The PP generation devices convert solar energy into electrical energy (Bawazir et al., 2023). The power generation efficiency of PP generation devices is related to the light intensity and temperature (Li et al., 2024). The output of power generation devices is expressed as follows (Equation 2).

$$P_{\mathrm{PV}} = P_{\mathrm{STC}} \frac{G_{\mathrm{ING}}}{G_{\mathrm{STC}}} \left[1 + k\left(T_C - T_r\right)\right] \tag{2}$$

where, P_{PV} is the output of PP generation devices; P_{STC} is the maximum output of the PP generation devices; G_{ING} is the intensity of light; G_{STC} is the standard light intensity; k is the temperature coefficient; T_C is the ambient temperature; T_r is the reference temperature.

2.3 Mathematical modeling of energy storage devices

The ES devices can absorb or release AP. When WP generation devices and PP generation devices generate more power than the load demand, ES devices can absorb power to avoid wasting electricity (Rostamnezhad et al., 2022). When the output power of WP devices and PP devices is less than the load demand, ES devices can release power to relieve the power tension (Song et al., 2023). Batteries are common ES devices. The most widely applied equivalent model for ES plants is the Davignan equivalent model. An ES device can be represented mathematically as follows (Equation 3).

$$SOC(t) = \begin{cases} SOC(t-1) + \dfrac{\eta_{\mathrm{ch}} I_t}{C_N} & \text{Charge} \\ SOC(t) - \dfrac{I_t}{C_N \eta_{dis}} & \text{Discharge} \end{cases} \tag{3}$$

where, $SOC(t)$ is the state of charge at time t; $SOC(t-1)$ is the state of charge of the ES device at time $t-1$; η_{ch} is the charge efficiency; η_{dis} is the discharge efficiency; C_N is the rated power; I_t is the current flows through ES devices.

3 Active power balance control method based on transfer learning double deep Q-network approach

This study proposes a TLDDQN-based AP balance control strategy. This AP balance control strategy based on TLDDQN is applied to cooperate with the traditional thermal generating units for AP balance control of the RE generation system by controlling storage devices in the WPS power system. The transfer learning method is employed to enhance the DDQN, thereby facilitating the formation of the TLDDQN. In addition, this study proposes an improved adaptive entropy mechanism to improve the exploratory ability of agents during the training process. The TLDDQN has the advantage of being able to adapt to different environments and can provide a strategy to maximize the cooperation of the WP and PP systems with the conventional units for the AP balance control of the renewable power system.

3.1 Transfer learning method

Transfer learning achieves the purpose of learning new knowledge quickly through the transfer of similarities (Wang et al., 2023). In contrast to traditional machine learning, transfer learning permits a relaxation of the fundamental assumption that the training data must independently satisfy the same distributional conditions as the test data. When training and test data have different distributions, transfer learning methods allow for fast model building.

The transfer learning approach defines a source domain D_s and a target domain D_t. The source and target domains have different data distributions $P(X_s)$ and $P(X_t)$. The focus of the transfer learning approach is finding the similarities between the source domains and target domains and utilize appropriately.

3.2 Double deep Q-network approach

The DQN employs a combination of deep learning methodologies and Q-learning to address the issue of dimensionality explosion that is inherent to the latter (Yi et al., 2022). The DQN algorithm applies NNs as function approximators to approximate the state-action value. The expression of the objective function of the DQN algorithm is expressed as follows (Equation 4).

$$Y_i^{\mathrm{DQN}} = r + \gamma \max_{a'} Q\left(s', a'; \theta'\right) \tag{4}$$

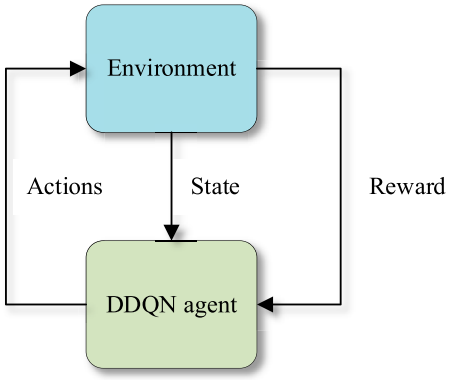

FIGURE 1
Relationship between DDQN agents and environments.

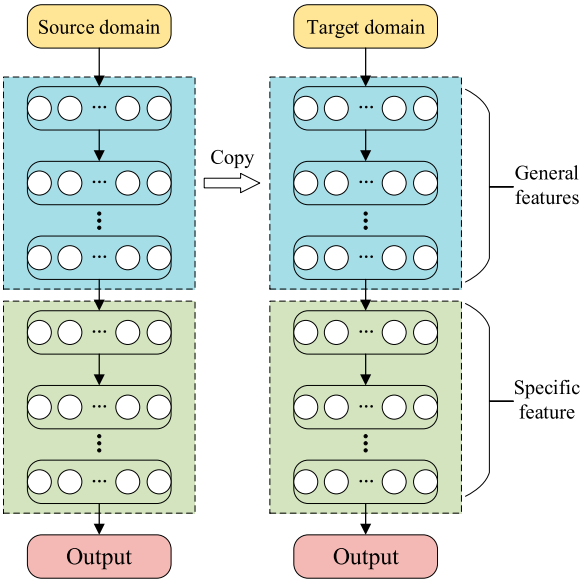

FIGURE 2
Transfer learning methods transfer NNs.

where, r is the reward of actions; γ is the discount factor; (s', a') is the state-action value at the next moment; θ' is the weight of the target network; max is taking the maximum value.

The DQN algorithm is susceptible to overestimation of the Q value. The DDQN algorithm represents an improvement from the original DQN algorithm. The DDQN algorithm separates the action selection and action valuation processes of the DQN algorithm, thus addressing the issue of the DQN algorithm being prone to overestimating the Q value. The optimization function Y_i^{DDQN} of the NN of the DDQN algorithm is expressed as follows (Equation 5).

$$Y_i^{\mathrm{DDQN}} = r + \gamma Q\left(s', \arg\max_{a'} Q(s', a'; \theta^-); \theta'\right) \tag{5}$$

where, the Q-function with weights θ' is applied to select the action behavior; the Q-function with weights θ^- is applied to evaluate the action.

Figure 1 illustrates the relationship between the DDQN agent and environments. The DDQN agent outputs actions to act on environments. The DDQN agent receives the reward value and state of the output actions from environments to update the parameters of agents.

3.3 Improvement of adaptive entropy mechanism

Ordinary reinforcement learning algorithms tend to converge to a local optimum solution in the late stage of training. To solve this problem, some reinforcement learning algorithms combine the entropy maximization method with the reinforcement learning algorithm to obtain stronger algorithmic performance. Reinforcement learning methods that combine the entropy of a policy to maximize the reward also maximize the entropy of the distribution of the actions of policy in each state, rather than just considering maximizing the reward of actions. As a result, compared with ordinary reinforcement learning methods, the reinforcement learning method with the entropy of policies obtains stronger exploration ability and effectively solves the problem of convergence to locally optimal solutions. Accordingly, this study combines the adaptive entropy mechanism into the DDQN algorithm and improves the adaptive entropy mechanism.

The entropy of a strategy is a measure of the uncertainty of a probability distribution. As the distribution becomes more random, the entropy value increases. Reinforcement learning algorithms combining the method of maximizing entropy for the augmentation and generalization of the rewards of agents can be expressed as follows (Equation 6).

$$r(s_t, a_t) = r(s_t, a_t) + \delta H(p) \tag{6}$$

where, $r(s_t, a_t)$ is the reward of the intelligent; δ is the adaptive entropy temperature coefficient; $H(p)$ is the entropy of the strategy.

According to the knowledge of information theory, the entropy of the strategy can be expressed as follows (Equation 7).

$$H(p) = -\sum_i p_i \log(p_i) \tag{7}$$

where, p_i is the state transfer distribution.

In the above process, the value of the adaptive entropy temperature coefficient is very important. Too small an adaptive entropy temperature coefficient will result in the agent easily converging to the local optimal solution; too large an adaptive entropy temperature coefficient will result in the agent generating too much unnecessary exploration. However, previous deep reinforcement learning algorithms do not provide reasonable values for the adaptive entropy temperature coefficient. Therefore, this study proposes improved adaptive entropy temperature coefficients to enhance the rationality of entropy utilization.

This study proposes a method to dynamically adjust the entropy temperature coefficient based on the average reward. If the average reward of an agent is stagnant or decreasing, the entropy value should increase to encourage the exploration of new strategies; on

the contrary, if the average reward continues to increase, the entropy value should decrease to stabilize the currently effective strategies. Therefore, the entropy temperature coefficient proposed in this study can be expressed as follows (Equation 8).

$$\delta' = \begin{cases} \delta'_{\max} \exp(-ite/50)\delta, & if\ (R_t > R_{t-1}) \\ \delta'_{\min}\delta, & else \end{cases} \tag{8}$$

where, δ' is the entropy temperature coefficient proposed in this study; R_t is the average reward of the intelligences; *ite* is the number of iterations in the training process; $\delta'_{\max}$ is the maximum entropy temperature coefficient; $\delta'_{\min}$ is the minimum entropy temperature coefficient; exp is the exponential operator.

3.4 Transfer learning double deep Q-network approach

This study proposes TLDDQN that is formed by the transfer learning approach combined into the DDQN approach. As shown in Figure 2, the NN of the DDQN approach can be split into two parts. One part of the NN is responsible for learning generic features. The other part of the double NN is responsible for learning task-specific features. First, when the deep reinforcement learning agents are under a new environment, the NN in the source domain that is responsible for learning generalized features is directly copied to the target domain. Besides, the corresponding NN parameters are frozen. Then, the transfer learning method randomly initializes the unfrozen NN parameters in the target domain and retrains NN parameters with the data in the target domain.

3.5 Transfer learning double deep Q-network-based active power balance control method for wind-photovoltaic-storage power systems

This study applies the proposed TLDDQN to control ES devices to fully consider the cost factor at the same time as the traditional unit to carry out AP balance control of WPS power systems. Considering the environmentally friendly and renewable advantages of wind and PP generation systems, the AP balance control strategy based on the proposed TLDDQN prioritizes the consumption of power generated by WP and PP generation systems. However, because of the stochastic and fluctuating characteristics of WP and PP generation systems, the power output of the WP-PP systems alone is challenged to match the load consumption. Therefore, the AP balance control strategy in this study applies the proposed TLDDQN method to control ES devices, which are combined with the traditional thermal power generation system for the AP balance control of the WPS power system.

The TLDDQN method is a deep reinforcement learning method that necessitates the definition of the state, action, and reward settings.

The state of an agent is the mathematical representation of the environment in which the agent is located. Therefore, in this study, the state of the agent includes the load power, the power generated by the wind power generator, the power generated by the photovoltaic power generator, and the charge state of the energy storage device at the same moment. Therefore, the state S_t of the agent can be represented as follows (Equation 9).

$$S_t = \{P_{\text{load}}, P_{\text{wt}}, P_{\text{pv}}, soc\} \tag{9}$$

where, P_{load} is the load power; *soc* is the battery status.

The action of the TLDDQN consists of a series of discrete variables. The action a_t is represented as follows (Equation 10).

$$a_t = \left\{ l, l + \frac{h-l}{M}, ..., h \right\} \tag{10}$$

where, l is the lower limit of the action value; h is the upper limit of the action value; M is the dimension of the action space.

The reward setting of the TLDDQN agent mainly takes into account the operational cost of the WPS power system and the discharge power of the ES device. The reward setting *rew* is expressed as follows (Equations 11–17).

$$rew = \alpha r_1(t) + \beta r_2(t) \tag{11}$$

$$r_1(t) = C_f(t) + C_{OM}(t) + C_{DEP}(t) + C_L \tag{12}$$

$$r_2(t) = P_{dis}(t) \tag{13}$$

$$C_f(t) = \sum_{i=1}^{N} C_{fuel} \frac{1}{LHV} \sum_{t=1}^{T} \frac{P_i(t)}{\eta_i(t)} \tag{14}$$

$$C_{OM}(t) = \sum_{i=1}^{N} K_{OM,i} P_i(t) \tag{15}$$

$$C_{DEP}(t) = \sum_{i=1}^{N} \frac{C_{ACC,i}}{8760 P_{ri} f_{cf,i}} P_i(t) \tag{16}$$

$$C_L = C_{\text{bu}} load \tag{17}$$

where, $r_1(t)$ is the WPS power system's operational cost reward; $r_2(t)$ is the ES unit's discharge power reward; α is the operating cost coefficient; β is the discharge power coefficient; N is the times that the AP balance control method is dispatched within a day; $P_{dis}(t)$ is ES unit's discharge power; $C_f(t)$ is the fuel cost consumed; $C_{OM}(t)$ is the maintenance cost; $C_{DEP}(t)$ is the depreciation cost; C_L is the compensation cost for the outage when the load is removed; C_{fuel} is the price of fuel; *LHV* is the low calorific value; $P_i(t)$ is the AP output of the generating unit; $\eta_i(t)$ is the fuel combustion efficiency of the thermal generating unit; $K_{OM,i}$ is the maintenance factor of the generating unit; $C_{ACC,i}$ is the installation cost of the generating unit; P_{ri} is the rated power of the generating unit; $f_{cf,i}$ is the capacity factor; C_{bu} is the compensatory price per unit of electricity; *load* is the excised amount of electricity.

Figure 3 shows the structure of the AP balance control method based on the proposed TLDDQN. The RE unit relies on wind and solar energy to generate electricity. The ES control center receives the power generation information of RE units, the load information, and the charge state information of ES devices. The control method of the ES control center is the AP balance control method based on the TLDDQN. The thermal power unit formulates the thermal power generation strategy based on the power generation situation of the ES device, the power generation situation of RE units, and the load power situation. Figure 4 shows the flowchart of the AP regulation of this study. When the power generated by a RE generator is greater than the load demand, ES device absorb as much of the excess power as possible. When the power generated by the RE

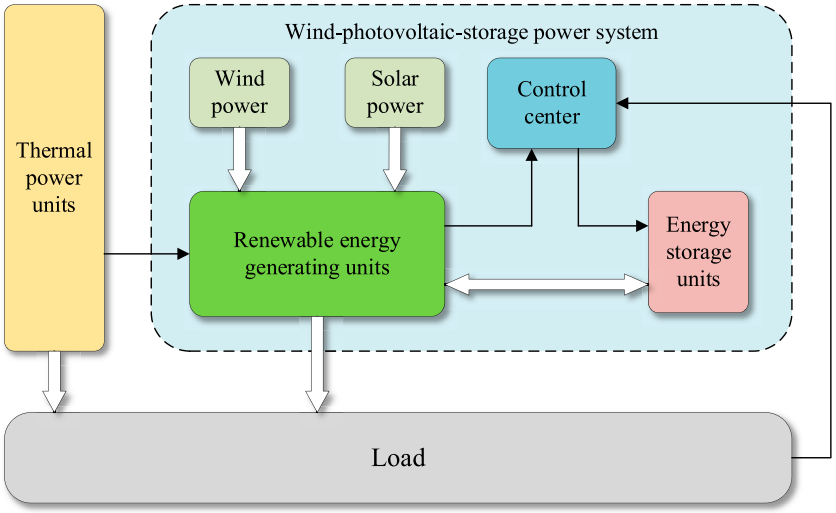

FIGURE 3
Structure of AP balance control method for WPS power system.

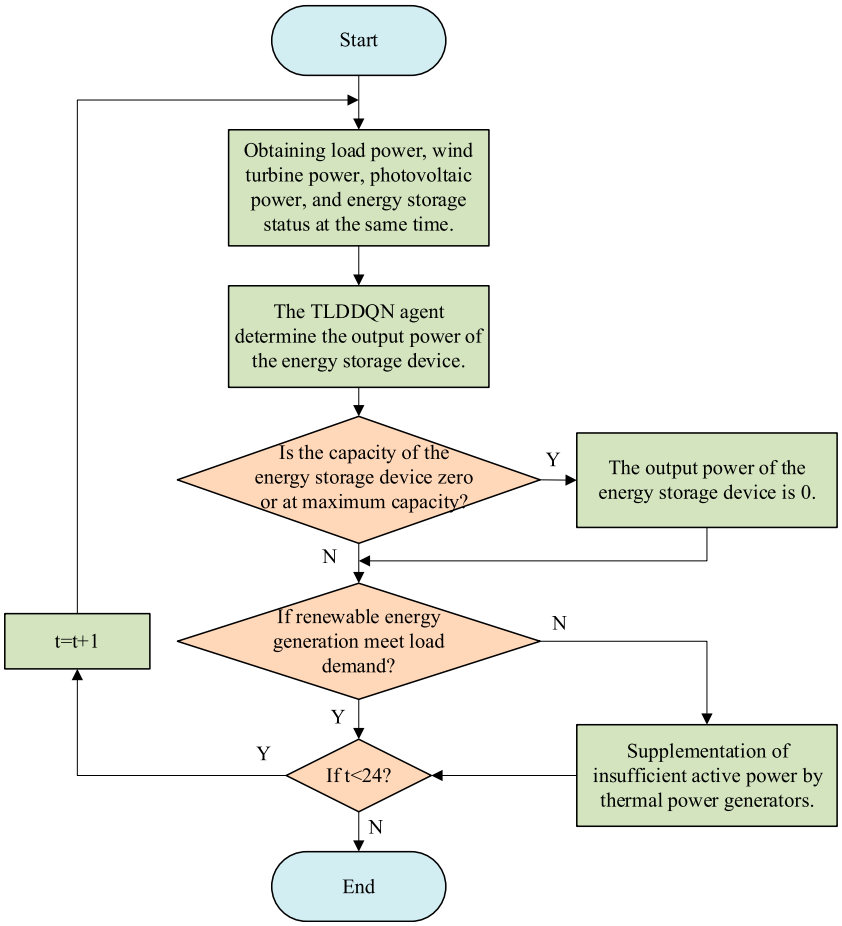

FIGURE 4
AP regulation flowchart.

10.3389/fenrg.2024.1448046

TABLE 1 Parameters of algorithms.

Algorithm	Paraments	Value
particle swarm optimization	Number of individuals in the population	30
particle swarm optimization	Number of iterations	500
TLDDQN	Greed rate	0.2
TLDDQN	Learning rate	0.05
TLDDQN	Power at the initial moment of the ES device	10
TLDDQN	Maximum capacity of the ES device	20
TLDDQN	Self-discharge rate of ES devices	0.001
TLDDQN	Maintenance costs of ES devices	0.0012
TLDDQN	Maximum output of gas turbines	65

TABLE 2 Load power, wind turbine power, and PP at a site for 24 h.

Time(h)	P_{load} (kW)	P_{wt} (kW)	P_{pv} (kW)
1	62.99	32.76	0
2	57.95	37.8	0
3	55.43	34.57	0
4	51.65	39.06	0
5	50.39	33.28	0
6	62.99	15.12	0
7	81.89	17.64	2.52
8	97.01	21.42	10.08
9	117.17	25.2	16.38
10	127.24	13.12	25.2
11	123.47	7.56	28.98
12	108.35	15.34	30.24
13	99.53	6.3	28.98
14	93.23	7.43	26.46
15	97.01	17.64	21.42
16	109.61	19.42	16.38
17	118.4	31.43	11.34
18	122.21	34.02	2.52
19	129.76	22.68	0
20	136.06	18.9	0
21	125.98	20.16	0
22	105.83	18.57	0
23	89.45	30.24	0
24	66.77	39.06	0

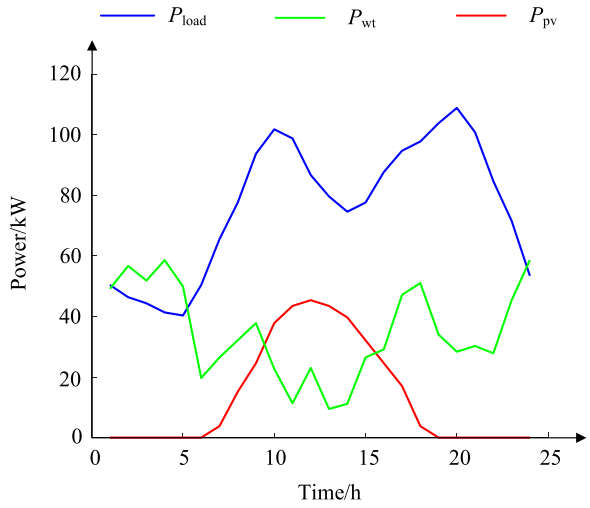

FIGURE 5
Curves of load power, turbine power, and PP at a site.

generator is less than the load demand, the ES device generates active power to reduce the power generated by the thermal generator.

4 Case studies

In this study, experiments are carried out to verify the effectiveness of the AP balance control method based on the TLDDQN proposed in this study based on load power, wind turbine power, and PP data at a site. This study compares the number of iterations required to accomplish convergence between the proposed TLDDQN and DDQN and the output of thermal power generation units by applying the proposed TLDDQN algorithm and particle swarm optimization for AP balance control of WPS power systems.

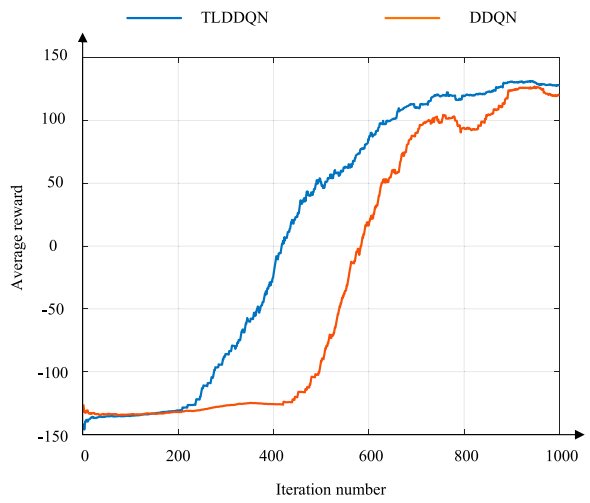

FIGURE 6
Average reward curve during training.

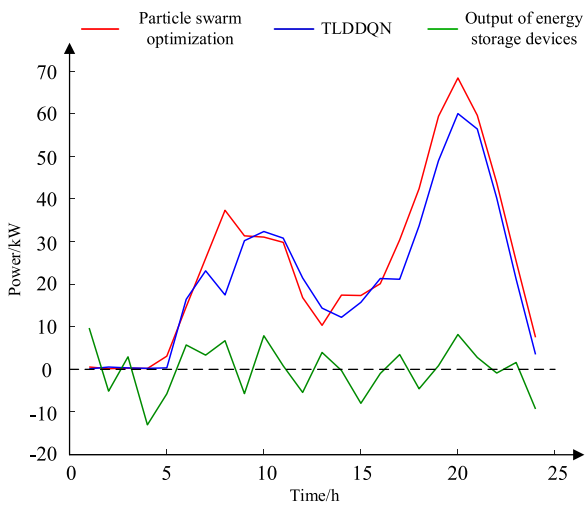

FIGURE 7
Output power of thermal power generators.

4.1 Experimental environment

The simulation software applied in this study is MATLAB R2023a. The simulations in this study were run on a personal computer with the operating system Windows 10, running memory of 16 GB, CPU model AMD R5 3600 (3.6 GHZ), and graphic processing unit model NVIDIA RTX 2070.

Table 1 shows the parameters of the algorithms involved in this study. Table 2 shows the load power, wind turbine power, and PP data updated hourly during a day at a site. Figure 5 shows the graphs of load power, turbine power, and PP obtained from the data in Table 1. Where, P_{load} is load power; P_{wt} is turbine power; P_{pv} is PP. The load power is low at night and high during the day. The wind turbine's power generation shows a large fluctuation during the day. The PP generation unit can only obtain power during the daytime resulting in a pronounced peak in the generation power curve.

4.2 Comparison of training processes

To verify the effectiveness of the TLDDQN algorithm proposed in this study in improving the convergence speed of agents. In this study, TLDDQN algorithm and DDQN algorithm are applied to train agents respectively.

Figure 6 shows the average reward curves of the TLDDQN algorithm and DDQN algorithm. Compared with the traditional DDQN algorithm, the TLDDQN algorithm proposed in this study introduces the adaptive entropy mechanism and makes improvements to the adaptive entropy mechanism. The introduction of the improved adaptive entropy mechanism can improve the exploratory ability of the agents during the training process. In addition, the TLDDQN algorithm proposed in this study introduces the TL method to improve the adaptability of agents. Therefore, compared with the traditional DDQN algorithm, the TLDDQN algorithm proposed in this study has stronger algorithmic performance. In the same environment, the number of iterations required for the TLDDQN agents proposed in this study to reach convergence is about 685. The number of iterations required for the DDQN agents to reach convergence is about 852. Compared to the traditional DDQN algorithm, the TLDDQN method reduces the training time by 19.60%.

In summary, the TLDDQN proposed in this study can converge faster than the traditional DDQN.

4.3 Comparison of adjustment effect

In this study, the AP balance control methods based on the proposed TLDDQN and the particle swarm optimization are applied to control ES devices in the experimental environment shown in Section 4.1, respectively.

The advantageous attributes of our proposed method, characterized by the TLDDQN, are encapsulated in its enhanced capability to modulate energy storage device outputs with precision, effectively addressing the intermittency of renewable energy sources and consequently leading to a substantial reduction in the operational burden on thermal power generation units. Figure 7 shows the thermal power generation power curves of the AP balance control method based on TLDDQN and the thermal power generation power curves of the particle swarm algorithm based on the particle swarm optimization. The AP balance control method based on TLDDQN reduces fossil energy consumption by 12.01% as compared to the particle swarm optimization-based AP balance control method.

In summary, the AP balance control method based on the proposed TLDDQN can solve the cooperation problem between the RE generation system and the traditional thermal generating units.

5 Conclusion

Aiming at the problem that thermal power generation units need to cooperate with RE generation units for the AP

balance control of the WPS power system when the proportion of RE generation devices is high, this study proposes the TLDDQN algorithm-based AP balance control method for the WPS power system. The proposed TLDDQN algorithm-based AP balance control method of the WPS power system can control the ES device of the WPS power system to balance the AP of the regional WPS power system. The features of the proposed AP balance control method for WPS power systems based on the TLDDQN algorithm are summarized as follows.

(1) The AP balance control method for WPS power systems based on the proposed TLDDQN algorithm can reduce the output of thermal power generators compared with the particle swarm optimization.
(2) The AP balance control method of the WPS system based on the proposed TLDDQN combines the advantages of fast learning possessed by transfer learning and the advantages of dealing with complex environments possessed by the DDQN algorithm. In addition, the improved adaptive entropy mechanism can improve the exploratory ability of agents during the training process. Therefore, the AP balance control method of the WPS system based on the proposed TLDDQN can precisely control the AP balance of the WPS system.

In future works, i) more types of RE generation units will be considered; ii) the proposed TLDDQN algorithm will be improved to increase the accuracy of power control.

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

JX: Supervision, Validation, Writing–original draft. WZ: Formal Analysis, Investigation, Writing–original draft. WL: Resources, Visualization, Writing–review and editing. YZ: Validation, Writing–review and editing. YoL: Conceptualization, Software, Writing–original draft. XM: Methodology, Writing–review and editing. YuL: Formal Analysis, Writing–review and editing.

Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. This research was funded by the State Grid Shanxi Electric Power Company Science and Technology Program, grant number 5205J0230001.

Conflict of interest

Authors JX, WZ, WL, YZ, YoL, XM, and YuL were employed by State Grid Shanxi Power Company Lvliang Power Supply Company.

The authors declare that this study received funding from State Grid Shanxi Electric Power Company. The funder had the following involvement in the study: study design, data collection and analysis, decision to publish, and preparation of the manuscript.

Publisher's note

References

Abdelghany, M. B., Al-Durra, A., Daming, Z., and Gao, F. (2024). Optimal multi-layer economical schedule for coordinated multiple mode operation of wind–solar microgrids with hybrid energy storage systems. *J. Power Sources* 591, 233844. doi:10.1016/j.jpowsour.2023.233844

Bawazir, R. O., Çetin, N. S., and Fadel, W. (2023). Optimum PV distributed generation based on grid and geographical area: a case study of Aden governorate, Yemen. *Energy Convers. Manag.* 297, 117703. doi:10.1016/j.enconman.2023.117703

Cheng, L., Liu, G., Huang, H., Wang, X., Chen, Y., Zhang, J., et al. (2020). Equilibrium analysis of general N-population multi-strategy games for generation-side long-term bidding: an evolutionary game perspective. *J. Clean. Prod.* 276, 124123. doi:10.1016/j.jclepro.2020.124123

Cheng, L., and Yu, T. (2019). A new generation of AI: a review and perspective on machine learning technologies applied to smart energy and electric power systems. *Int. J. Energy Res.* 43 (6), 1928–1973. doi:10.1002/er.4333

Dong, H., Fu, Y., Jia, Q., and Wen, X. (2022). Optimal dispatch of integrated energy microgrid considering hybrid structured electric-thermal energy storage. *Renew. Energy* 199, 628–639. doi:10.1016/j.renene.2022.09.027

Grover, H., Verma, A., and Bhatti, T. S. (2022). DOBC-based frequency and voltage regulation strategy for PV-diesel hybrid microgrid during islanding conditions. *Renew. Energy* 196, 883–900. doi:10.1016/j.renene.2022.06.140

Guerra, K., Haro, P., Gutiérrez, R. E., and Gómez-Barea, A. (2022). Facing the high share of variable renewable energy in the power system: Flexibility and stability requirements. *Appl. Energy* 310, 118561. doi:10.1016/j.apenergy.2022.118561

Han, Y., Liao, Y., Ma, X., Guo, X., Li, G., and Liu, X. (2023). Analysis and prediction of the penetration of renewable energy in power systems using artificial neural network. *Renew. Energy* 215, 118914. doi:10.1016/j.renene.2023.118914

Hao, X., Chen, Y., Wang, H., Wang, H., Meng, Y., and Gu, Q. (2023). A V2G-oriented reinforcement learning framework and empirical study for heterogeneous electric vehicle charging management. *Sustain. Cities Soc.* 89, 104345. doi:10.1016/j.scs.2022.104345

Hasanien, H. M., Alsaleh, I., Tostado-Véliz, M., Zhang, M., Alateeq, A., Jurado, F., et al. (2024). Hybrid particle swarm and sea horse optimization algorithm-based optimal reactive power dispatch of power systems comprising electric vehicles. *Energy* 286, 129583. doi:10.1016/j.energy.2023.129583

Jiang, B., Lei, H., Li, W., and Wang, R. (2022). A novel multi-objective evolutionary algorithm for hybrid renewable energy system design. *Swarm Evol. Comput.* 75, 101186. doi:10.1016/j.swevo.2022.101186

Jiang, W., Liu, Y., Fang, G., and Ding, Z. (2023). Research on short-term optimal scheduling of hydro-wind-solar multi-energy power system based on deep reinforcement learning. *J. Clean. Prod.* 385, 135704. doi:10.1016/j.jclepro.2022.135704

Jung, C., and Schindler, D. (2023). The properties of the global offshore wind turbine fleet. *Renew. Sustain. Energy Rev.* 186, 113667. doi:10.1016/j.rser.2023.113667

Li, K., Ye, N., Li, S., Wang, H., and Zhang, C. (2023). Distributed collaborative operation strategies in multi-agent integrated energy system considering integrated demand response based on game theory. *Energy* 273, 127137. doi:10.1016/j.energy.2023.127137

Li, S., Deng, N., Lee, X., Yan, S., and Chen, C. (2024). Optimal configuration of photovoltaic microgrid with improved ant colony dynamic programming. *J. Energy Storage* 83, 110714. doi:10.1016/j.est.2024.110714

Liu, Y., and Wang, J. (2022). Transfer learning based multi-layer extreme learning machine for probabilistic wind power forecasting. *Appl. Energy* 312, 118729. doi:10.1016/j.apenergy.2022.118729

Mehmood, A., Raja, M. A. Z., and Jalili, M. (2023). Optimization of integrated load dispatch in multi-fueled renewable rich power systems using fractal firefly algorithm. *Energy* 278, 127792. doi:10.1016/j.energy.2023.127792

Moosavian, S. F., Noorollahi, Y., and Shoaei, M. (2024). Renewable energy resources utilization planning for sustainable energy system development on a stand-alone island. *J. Clean. Prod.* 439, 140892. doi:10.1016/j.jclepro.2024.140892

Revathi, R., Senthilnathan, N., and V, K. C. (2024). Hybrid optimization approach for power scheduling with PV-battery system in smart grids. *Energy* 290, 130051. doi:10.1016/j.energy.2023.130051

Rostamnezhad, Z., Mary, N., Dessaint, L. A., and Monfet, D. (2022). Electricity consumption optimization using thermal and battery energy storage systems in buildings. *IEEE Trans. Smart Grid* 14 (1), 251–265. doi:10.1109/tsg.2022.3194815

Russo, M. A., Carvalho, D., Martins, N., and Monteiro, A. (2023). Future perspectives for wind and solar electricity production under high-resolution climate change scenarios. *J. Clean. Prod.* 404, 136997. doi:10.1016/j.jclepro.2023.136997

Song, Y., Mu, H., Li, N., Wang, H., and Kong, X. (2023). Optimal scheduling of zero-carbon integrated energy system considering long-and short-term energy storages, demand response, and uncertainty. *J. Clean. Prod.* 435, 140393. doi:10.1016/j.jclepro.2023.140393

Wang, K., Wang, H., Yang, Z., Feng, J., Li, Y., Yang, J., et al. (2023). A transfer learning method for electric vehicles charging strategy based on deep reinforcement learning. *Appl. Energy* 343, 121186. doi:10.1016/j.apenergy.2023.121186

Yakout, A. H., Hasanien, H. M., Turky, R. A., and Abu-Elanien, A. E. (2023). Improved reinforcement learning strategy of energy storage units for frequency control of hybrid power systems. *J. Energy Storage* 72, 108248. doi:10.1016/j.est.2023.108248

Ye, L., Jin, Y., Wang, K., Chen, W., Wang, F., and Dai, B. (2023). A multi-area intra-day dispatch strategy for power systems under high share of renewable energy with power support capacity assessment. *Appl. Energy* 351, 121866. doi:10.1016/j.apenergy.2023.121866

Yi, Z., Luo, Y., Westover, T., Katikaneni, S., Ponkiya, B., Sah, S., et al. (2022). Deep reinforcement learning based optimization for a tightly coupled nuclear renewable integrated energy system. *Appl. Energy* 328, 120113. doi:10.1016/j.apenergy.2022.120113

Yin, L., and Wu, Y. (2022). Mode-decomposition memory reinforcement network strategy for smart generation control in multi-area power systems containing renewable energy. *Appl. Energy* 307, 118266. doi:10.1016/j.apenergy.2021.118266

Yin, L., and Zheng, D. (2024). Decomposition prediction fractional-order PID reinforcement learning for short-term smart generation control of integrated energy systems. *Appl. Energy* 355, 122246. doi:10.1016/j.apenergy.2023.122246

Zhang, B., Hu, W., Xu, X., Li, T., Zhang, Z., and Chen, Z. (2022). Physical-model-free intelligent energy management for a grid-connected hybrid wind-microturbine-PV-EV energy system via deep reinforcement learning approach. *Renew. Energy* 200, 433–448. doi:10.1016/j.renene.2022.09.125