frontiers | Frontiers in Energy Research

TYPE Original Research
PUBLISHED 14 September 2023
DOI 10.3389/fenrg.2023.1272216

OPEN ACCESS

EDITED BY
Shiwei Xie,
Fuzhou University, China

REVIEWED BY
Menglin Zhang,
China University of Geosciences Wuhan, China
Yachao Zhang,
Fuzhou University, China

*CORRESPONDENCE
Zhenning Pan,
scutpanzn@163.com

RECEIVED 03 August 2023
ACCEPTED 05 September 2023
PUBLISHED 14 September 2023

CITATION
Huang W, Dai Z, Hou J, Liang L, Chen Y, Chen Z and Pan Z (2023), Risk-averse stochastic dynamic power dispatch based on deep reinforcement learning with risk-oriented Graph-Gan sampling. *Front. Energy Res.* 11:1272216. doi: 10.3389/fenrg.2023.1272216

Risk-averse stochastic dynamic power dispatch based on deep reinforcement learning with risk-oriented Graph-Gan sampling

Wenqi Huang[1], Zhen Dai[1], Jiaxuan Hou[1], Lingyu Liang[1], Yiping Chen[1], Zhiwei Chen[2] and Zhenning Pan[2]*

[1]Digital Grid Research Institute, China Southern Power Grid, Guangzhou, Guangdong, China, [2]School of Electric Power Engineering, South China University of Technology, Guangzhou, Guangdong, China

The increasing penetration of renewable energy sources (RES) brings volatile stochasticity, which significantly challenge the optimal dispatch of power systems. This paper aims at developing a cost-effective and robust policy for stochastic dynamic optimization of power systems, which improves the economy as well as avoiding the risk of high costs in some critical scenarios with small probability. However, it is hard for existing risk-neutral methods to incorporate risk measure since most samples are normal. For this regard, a novel risk-averse policy learning approach based on deep reinforcement learning with risk-oriented sampling is proposed. Firstly, a generative adversarial network (GAN) with graph convolutional neural network (GCN) is proposed to learn from historical data and achieve risk-oriented sampling. Specifically, system state is modelled as graph data and GCN is employed to capture the underlying correlation of the uncertainty corresponding to the system topology. Risk knowledge is the embedded to encourage more critical scenarios are sampled while aligning with historical data distributions. Secondly, a modified deep reinforcement learning (DRL) with risk-measure under soft actor critic framework is proposed to learn the optimal dispatch policy from sampling data. Compared with the traditional deep reinforcement learning which is risk-neutral, the proposed method is more robust and adaptable to uncertainties. Comparative simulations verify the effectiveness of the proposed method.

KEYWORDS

optimal dispatch, deep reinforcement learning, risk-oriented sampling, risk knowledge embedded, graph convolutional networks

1 Introduction

With the rapid development of power electronics, it is foreseeable that the proportion of renewable energy sources (RES) in the power system will continue to increase (Mathiesen et al., 2015). On the one hand, utilizing RES in future smart grids can help energy systems cope with energy depletion crisis. On the other hand, the uncertainty brought by RES makes the scheduling decision of the power system a greater security risk (Zhang et al., 2021). These challenges have a profound impact on the reliability and economy of power grid operations. Consequently, finding effective and reliable dispatch decisions has become a critical scientific

challenge with direct implications for operational safety. Dynamic economic dispatch (DED) is a dispatch strategy which allows dispatch decision to be given and adapted in response to the realizations of uncertainties evolutions. In this regard, developing the optimal policy (policy is a function about how system operator makes dispatch decision) for stochastic dynamic dispatch is crucial to maintain the supply-demand balance for power system under high renewable energy penetration. However, accurately modeling and solving the DED are required to address the challenges associated with uncertainties which attract substantial interest from both the electricity industry and academia.

Theoretically, stochastic dynamic dispatch of power systems is a typical multistage sequential decision problem. It usually contains enormous state and action space, and complex uncertainty variations, which makes its optimal policy almost intractable. In the past decade, extensive studies have devoted to developing the optimal policy for stochastic dynamic dispatch, mainly including look-ahead dispatch policy (also known as model predictive control), dynamic programming, and reinforcement learning. Among these methodologies, deep reinforcement learning (DRL) (Silver et al., 2016) is regarded as a promising alternative due to its strong nonlinear fitting ability, adaptability, and generalization. Through enough learning from training samples, its decision is adapted to the uncertainties observed overtime. Owing to these advantages, DRL has been widely applied to corresponding dynamic dispatch problems in smart grids. Reference (Hua et al., 2019) adopts a synchronous advantage actor-critic (A3C) to solve the energy management problem of continuous time coupling. Meanwhile Bedoya et al. (2021) solve an MDP problem considering the asynchronous data arrival using deep Q-network (DQN) and (Zhao and Wang, 2021) proposed an approach combining a GCN with a DQN to conduct sequential system restoration.

Although RL has been successfully applied in the optimal dispatch problem, most of them consider a risk-neutral objective. That is, they directly use original historical data as training samples and minimize the expectation of accumulative rewards. Such policy performs well in most scenarios or normal scenarios, however, when encountering some critical scenarios of which the possibility is small but the outcome is severe, e.g., network constraints violation or even supply-demand unbalance. The main difficulties to incorporate risk measurement can be summarized as follows: firstly, the critical scenarios with small possibility may be drown in massive normal scenarios, it is hard for algorithm to distinguish and learn these critical scenarios. Secondly, most RL methods use the average reward of batch samples for learning, risk measure is not considered. Some studies and our previous researches (Pan et al., 2020) have proposed a risk-averse RL for stochastic dynamic dispatch of multi-energy system, however, they directly used original historical data or Monte-Carlo sampling to form large batch of samples to compute risk adjusted objective. Note that the distribution of critical scenarios is quite sparse. Such approach cannot ensure critical scenarios with high cost and small probability are effectively sampled, leading to slow convergence and low sample efficiency. Reference (Liu et al., 2018) employs function approximation to avoid the trouble of stochastic modeling. Some literature simplify the problem by discretization, bringing the dilemma of inaccuracy and dimension disaster (Yu et al., 2015; Chen et al., 2019). Guo et al. deployed a novel policy-based PPO algorithm for a real-time dynamic optimal energy management in microgrids to make optimal scheduling decisions (Guo et al., 2022). Chen et al. developed a DDPG algorithm based on hybrid energy scheduling, which can learn the optimal policies from historical experiences, avoid inadequate exploration by introducing decaying noise (Chen et al., 2022). Reference (GUAN et al., 2020; Lv et al., 2020) has undertaken initial explorations into the utilization of deep reinforcement learning for real-time grid scheduling optimization. While these preliminary forays have delved into the optimization of grid scheduling, they have not yet been extended to address intricacies such as intra-day rolling scheduling, multi-objective grid scheduling, and the dynamic considerations arising from maintenance or minor faults in the system's topology. The above studies focus on simplified models for training RL and lack analysis and discussion of historical data.

Since RL can be regarded as a data-driven approach, its performance depends on the sampling data. Although a risk-averse or robust objective can be merged into traditional RL, another critical problem is how to ensure the risk scenarios with small probability are effectively sampled during learning? Since power system is mostly in a normal state, critical scenarios, e.g., line overloading, voltage violations, and load shedding unusually occurs. Existing methods directly use historical data as learning samples, however, this leads to slow convergence or invalid learning since critical scenarios are insufficient sampled.

To address the aforementioned key technical challenges, including the lack of risk-directed samples and the low robustness of policy, a novel risk-averse policy learning approach based on DRL with risk-oriented sampling is proposed. Firstly, a graph generative adversarial network (GGAN) that combines GANs (Goodfellow et al., 2014; Arjovsky and Bottou, 2017; Chen et al., 2018; Zhang et al., 2021) and GCNs (Shervashidze et al., 2009) is proposed. This integration allows to leverage historical graph data and capture the underlying correlation of the uncertainty corresponding to the system topology. Notably, GGAN incorporates risk knowledge to ensure that critical scenarios can be sufficiently generated while aligning with the underlying original data distribution. This modification boosts the interaction frequency between the agent and risk scenarios, enabling the identification and learning of crucial embeddings. Secondly, the existing DRL framework, specifically the SAC algorithm, is modified by incorporating risk measure. Consequently, the agent is incentivized to develop a cost-effective and robust policy for stochastic dynamic optimization, resulting in not only enhancement of the economy but also mitigating the risk of high costs in critical scenarios with low probabilities.

The specific contributions of this paper are as follows.

1) Risk-averse stochastic dynamic dispatch scheme: A DRL based risk-averse stochastic dynamic dispatch approach is proposed to enhance the robustness and economy of policy when encountering critical risk scenarios in power systems. To tackle the challenges of existing methods in inadaptability of risk measure and invalid sampling, this paper focuses on two aspects: data expansion and algorithm improvement. Specifically, firstly, risk-oriented sampling is proposed to

generate enough critical scenarios. Then, these samples are leaned by a risk measure incorporated DRL algorithms. By such way, the dispatch policy not only improves the economy but also avoid the risk of high costs in some rare but critical scenarios.

2) Risk-oriented sampling: to avoid the critical scenarios with small possibility being drown in massive normal scenarios, a risk-oriented sampling is proposed to generate more critical scenarios while maintaining the original data distribution. To achieve this, KEG_GAN (Knowledge Embedding Graph Generative Adversarial network) is proposed. Firstly, a graph representation is proposed to integrate node features with topology changes, allowing for the incorporation of topology information into the system state while achieving efficient expression of operational state. Secondly, through incorporating regularization terms into the loss function and leveraging topological connection relationship in the graph structure, the knowledge embedding guides data-driven model to generate risk-oriented samples. Thirdly, this paper proposed differentiated weighting method for batch samples with hierarchic stepped thresholds to enhance the utilization efficiency of critical samples.

2 Problem statement and proposed method

We first discuss the challenges in applying DRL for grid control under fast-changing power grid operation scenarios with increased uncertainties, which necessitates and highlights the need of risk control capability for DRL-based agents. Then, we introduce the framework of the method we proposed and how they solve the problem of optimal dispatch in power systems.

2.1 Problem statement and formulation

With the increased integration of RES into the power grid, ensuring economic efficiency in power system dispatching operations requires the consideration of operational risks in low-probability scenarios. While these risks may have a low likelihood of occurrence, their potential impact on the safety of the power grid cannot be underestimated.

The optimal dispatch problem entails learning a policy that enhances economic performance while mitigating the risk of incurring high costs in critical scenarios with small probabilities. Consequently, the dispatch problem in power systems, taking risk into account, can be represented by the following equation:

$$\begin{aligned} &obj1\ \min E\left[\sum_{t=0}^{t} C_r(x_t)\right] \\ &obj2\ \min E\left[\sum_{t=0}^{t} C_e(x_t)\right] \\ s.t.\quad &h_k(x_t) \le 0, k = 1, \ldots, m \\ &I_j(x_t) = 0, j = 1, \ldots, n \end{aligned} \tag{1}$$

Where $E(x)$ represents the expectation operator. $C_e(x)$ represents economic costs and $C_r(x)$ represents the cost of risk. $h_k(x)$ and $I_j(x)$ represent the physical constraints of the power system.

During power system operation, the primary objective is to guarantee the safety and reliability of the grid. Therefore, in addition to minimizing *obj*1 (e.g., risk considerations), *obj*2 (e.g., economic costs) should be taken-into-accounted as a secondary objective. The dispatch policy should prioritize minimizing *obj*1 while considering *obj*2 to ensure that the power system operates efficiently while maintaining a high level of safety and reliability.

The process of DRL solving the above problem can be defined as policy search in a Markov Decision Process (MDP) defined by a tuple (S, A, p, r, y), where S is the state space, A is the action space, p: $S\times A\rightarrow S$ is the transition function and p:$S\times A\rightarrow R$ is the reward function. The goal of DRL is to learn a policy $\pi_\theta(s_t)$:$S\rightarrow A$, such that it maximizes the expected accumulative reward $J(\pi_\theta)$ over time under p:

$$J(\pi_\theta) = E_{s_0,a_0,\ldots,s_t,s_t}\left[\sum_{t=0}^{t} y_i r(s_t, a_t)\right] \tag{2}$$

$$r(s_t, a_t) = f(C_e(x_t) + C_r(x_t)) \tag{3}$$

Where $a_t\sim\pi_\theta(s_t)$ and $s_{t+1}\sim p$ (s_t, a_t), and τ is the dispatch period. Note that maximizing the cumulative reward is the opposite of minimizing the cost, $f(x)$ achieving the conversion from cost to reward. The policy is parameterized by a neural network with weights θ in DRL. The traditional DRL framework is shown in the upper part of Figure 1.

There is a notable discrepancy in sample sizes across various scenarios, such as normal operation scenarios and high-risk operation scenarios or critical scenarios. During the agent's interaction with the environment, infrequent critical scenarios inundate the buffer, leading to policy updates that prioritize minimizing economic costs without adequately considering the security of power systems.

The uncertainty associated with RES presents a challenge for DRL, often leading to decisions that result in unsafe grid operations. Additionally, the uneven distribution of samples further compounds these issues, making it even more difficult to address the aforementioned challenges. To overcome these challenges, this paper proposes a method to enhance the DRL and effectively tackle these issues.

2.2 The basic framework of our method

The fundamental framework of the proposed risk-oriented Graph-Gan sampling assisted DRL for risk-averse stochastic dynamic dispatch, as well as the comparison with traditional DRL are illustrated in Figure 1.

The following improvements are made.

(1) Risk scenario generation: The sampling process for scenarios from the power system is modified to increase the proportion of risk scenarios while maintaining an appropriate balance with normal operation scenarios. This adjustment leads to a more risk-averse strategy, as depicted by the red circle in Figure 1.

(2) Risk probability sampling: To enhance the decision-making robustness of the intelligent agent, the importance of risk experience sampling is given higher priority during the update process. The policy is updated to ensure that the intelligent agent primarily learns from experiences related to

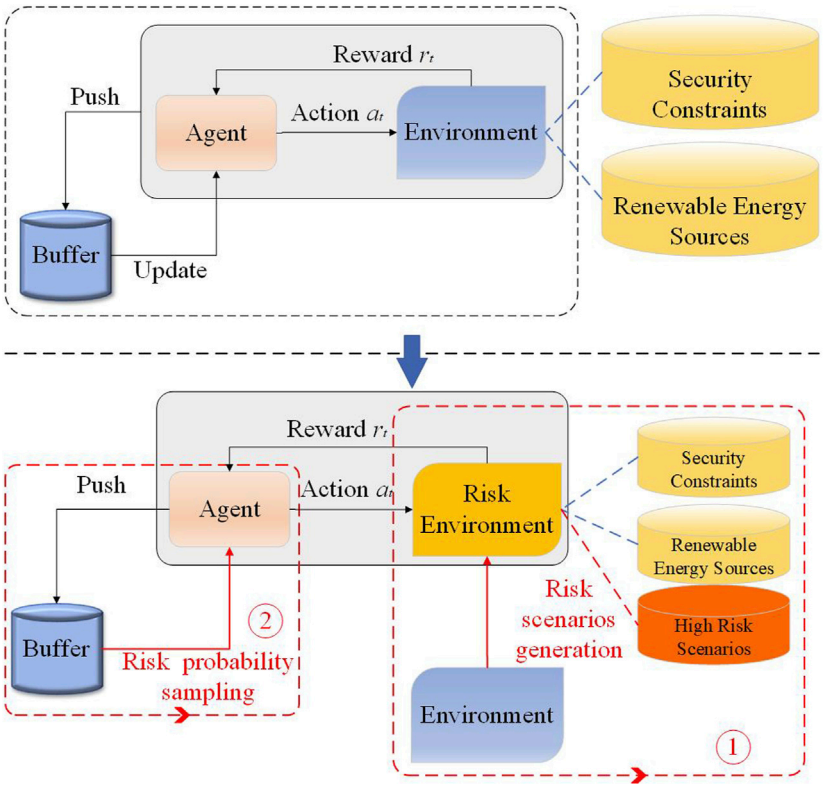

FIGURE 1
The framework of DRL with risk-oriented Graph-Gan sampling.

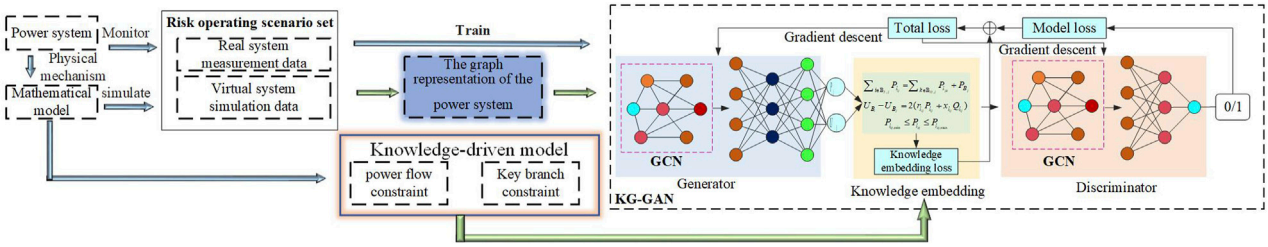

FIGURE 2
The framework of generation of extreme operation scenario sample based on KEG_GAN.

high-risk operation scenarios. This adjustment is visualized by the red circle in Figure 1.

3 Risk scenarios generation

The operation scenario data which is used to train the agent primarily originates from the measurement data of the actual power system. Critical scenarios, which often involve network constraints and can be mathematically described, usually occur very rarely in datasets. To address the issue of sparse data in the training scenario, data augmentation techniques can be employed to enhance the learning ability of the agent. However in the traditional data augmentation techniques such as GAN, only the path represented by the blue arrow in Figure 2 is considered. The model incorporates a deep understanding of the data distribution and is less inclined to generating outputs based on extremely scenarios during the generation process.

To address these limitations, this paper introduces KEG_GAN, which integrates risk knowledge and equations to effectively guide the training process. This approach leverages data-driven methods while incorporating additional guidance from risk knowledge, as depicted by the blue and green arrows in Figure 2. By combining data-driven learning with the integration of domain-specific knowledge, KEG_GAN aims to overcome the aforementioned challenges.

This framework begins by acquiring the operation scenario dataset and power grid topology information from measurement data and simulation systems within the power system. During the process of embedding knowledge into the model, the operation scenario data incorporates risk constraints. Drawing upon risk knowledge, the mechanism model of power grid operation scenes is formulated and analyzed, leading to the identification and extraction of risk equations present within the operation scenarios. These equations, such as power flow constraints and section constraints, are regularized and integrated into the GAN architecture to guide the model's learning process. In this paper, the operation scenarios are categorized into two groups: extreme operation scenarios with low safety margins and safe operation scenarios with high safety margins.

3.1 Feature extraction of power grid operation scene based on graph representation

In addition to node attributes and outputs, the power grid topology information plays a crucial role in capturing the key characteristics of operation scenarios. However, many conventional methods for generating scenarios focus solely on node-level data without considering the integration of power grid topology information. This limitation makes it difficult to generate operation scenarios that reflect the inherent coupling relationships between nodes using GAN-based approaches.

To address this limitation, effectively combining power grid topology information with operation scenario data becomes an essential approach to enhance the generation of power grid operation scenarios. In this paper, a graph representation is employed by combining node-level data of operation scenarios with power grid topology information. To effectively capture the information embedded in power grid operation scenarios, GCN are introduced to enhance the traditional GAN framework. This integration allows for the effective exploration and mining of critical information within power grid operation scenarios using GAN-based techniques.

The idea of GCN is to aggregate the information of neighbor nodes and obtain more powerful feature expression, which can dig deeply into the potential distribution of power grid operation scenario data. The calculation formula is shown as follows:

$$H^{(l+1)} = f\left(H^{(l)}, A\right) = \sigma\left(D^{-0.5}A'D^{-0.5}H^{(l)}W^{(l)}\right) \tag{4}$$

Where, $A' = A + I_N$ is an adjacency matrix with self-connection, I_N is the identity matrix. $D_{ii} = \sum_j A_{ij}'$ is the degree matrix of A'. $W^{(l)}$ is the trainable parameter in the convolutional layer of the GCN, and $H^{(l)}$ represents the input characteristics of the l layer. After matrix multiplication of the above formula, forward propagation is carried out through activation functions $\sigma(\cdot)$ such as RELU $(\cdot)$ (Thomas and Max, 2016).

The topological information is constructed and processed by GCN to mine the correlation between neighbor nodes and improve the learning effect of GAN. For this reason, the graph representation of the power grid operation scenario is shown in Figure 3. Firstly, sampling is conducted from the power system to obtain the operation scenario data of load, node voltage, the output of traditional unit and RES, etc. Then, the characteristic matrix H of the power grid operation scenario is constructed. The matrix size is N^*T, where T = {load, voltage., renewable energy output}, and the grid topology represents the connection relationship between nodes, which is represented by the adjacency matrix A. Therefore, the grid operation scenario is represented by the adjacency matrix A and the characteristic matrix H of the grid topology.

3.2 Knowledge embedding within agents considering operational risk

As mentioned previously, the data collected from the power system for generating operation scenarios includes various parameters such as active and reactive power of each node, node voltage, generator terminal voltage, and power output. However, when these data are not separated, it becomes challenging to extract the underlying physical constraints through data-driven methods alone. Considering the operational risks involved, it becomes necessary to incorporate human knowledge to uncover the physical mechanisms behind power grid operation scenarios and integrate them into the model to guide its learning process.

To address this, this paper introduces the concept of embedding knowledge into the model, with the goal of leveraging human knowledge to analyze and model the problem. This approach involves constructing mathematical equations that accurately represent the real physical situation. By incorporating a regularization term into the loss function, the mathematical equations derived from human knowledge are embedded into the neural network model, enabling guidance and modification of the data-driven model.

In line with the widespread concern regarding the risk of terminal voltage crossing the lower limit in power grid operation, this paper considers the scenario where the voltage at key nodes in the system approaches the critical lower limit as one of the critical scenarios. The physical constraints of this scenario are represented by Eqs 5–8, with the power flow constraints being expressed using simplified linear power flow equations (Baran and Wu, 1989).

$$\sum_{i\in B_{I,j}} P_{ij} = \sum_{k\in B_{O,j}} P_{jk} + P_j \tag{5}$$

$$\sum_{i\in B_{I,j}} Q_{ij} = \sum_{k\in B_{O,j}} Q_{jk} + Q_j \tag{6}$$

$$U_i - U_j = 2\left(r_{ij}P_{ij} + x_{ij}Q_{ij}\right) \tag{7}$$

$$U_{j,min} \le U_j \le U_{j,max} \tag{8}$$

Where, $B_{I,j}$,$B_{O,j}$ represent the node set that injects and flows node j along the reference direction. P_{ij},Q_{ij},P_j,Q_j represent active and reactive power of branch l_{ij} and node j. R_{ij} and x_{ij} are the resistance and reactance of the branch. The voltage of node j needs to meet its upper and lower limit constraints (8). $U_{\text{key,min}}$ is the lower limit of the voltage of the key node.

Simultaneously, the key section of the power system bears the significant responsibility of power transmission during grid operation. Thus, ensuring the reliability of electric energy delivered through the key section is a crucial task for the system's safe operation. However, the availability of extreme operation scenario where the power at the critical cross-section approaches the maximum transmission limit is relatively limited.

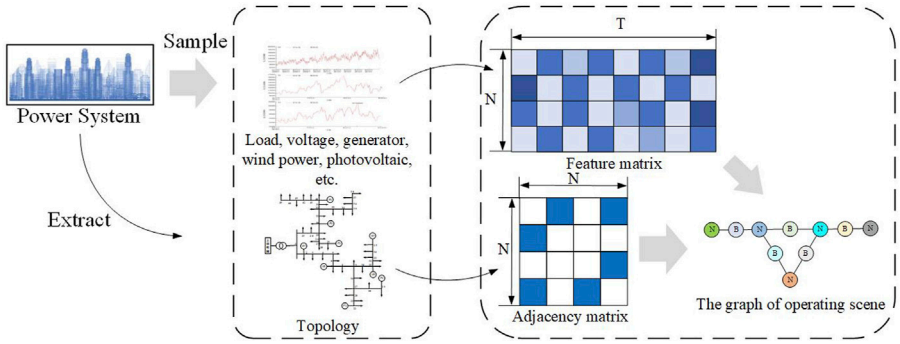

FIGURE 3
The graph representation of the power system operation scene.

Therefore, determining the critical upper limit for power transmission at the critical section of the power grid becomes a vital aspect of generating extreme operation scenarios. The corresponding constraints can be mathematically expressed through Eqs 9–10.

$$P_{ij,min} \le P_{ij} \le P_{ij,max} \tag{9}$$

$$Q_{ij,min} \le Q_{ij} \le Q_{ij,max} \tag{10}$$

Where, $P_{ij,\min}$,$P_{ij,\max}$,$Q_{ij,\min}$,$Q_{ij,\max}$ represent the maximum and minimum values of active power and reactive power that the tidal current section can flow through respectively.

Furthermore, the primary focus of this paper is on critical scenarios where the power flow in certain key branches, denoted as P_{key}, exceeds the upper limit threshold, posing a risk. To address this concern, equation constraint (11) is introduced, where $P_{\text{key,max}}$ represents the power flow upper limit of the key branches. Consequently, in the KEG_GAN framework, it is essential to ensure that the generated operation scenes comply with the aforementioned constraints to the best extent possible. The loss function of KEG_GAN can be formulated as follows:

$$L_{\text{total}} = L_{\text{model}} + L_{\text{constraint},i}, i = 1, 2 \tag{11}$$

Where, L_{model} represents the loss function of Graph-GAN, and $L_{\text{constraint},i}$ represents the loss function of the regularization knowledge embedding model with section constraint or voltage constraint. Therefore, in KEG_GAN, according to the training objectives of generator G and discriminator D, the sum of loss functions of the KEG_GAN are shown in Eqs 12, 13, respectively.

$$L_G = -E_{z\sim p_z(z)}[log(1 - D(G(z)))] \tag{12}$$

$$L_D = -E_{x\sim p_{\text{data}}(x)}[log D(x)] + E_{z\sim p_z(z)}[log(1 - D(G(z)))] \tag{13}$$

Where, E represents the distribution expectation of samples, p_{data}(x) represents the probability distribution of real sample x, and p_z(x) represents the probability distribution of generating sample z. Based on the above equation, the objective function of the adversarial network generated by Eq. 14 can be derived:

$$\begin{aligned} L_{\text{model}} &= \min_G \max_D W(D,G) \\ &= E_{x\sim p_{data}(x)}[log D(x)] \\ &\quad + E_{z\sim p_z(z)}[log(1 - D(G(z)))] \end{aligned} \tag{14}$$

Key section constraints are written into the model by means of loss function regularization, which can be expressed by Equation 15:

$$\begin{aligned} L_{\text{constraint},1} &= min\left(\frac{1}{N_l}\sum_{i=1}^{N_u}\left|P_{\text{key}} - P_{\text{key},max}\right|^2\right) \\ &= min\left(\text{MSE}_{P_{\text{key}},P_{\text{key},max}}\right) \end{aligned} \tag{15}$$

Where, N_l represents the number of key cross sections; MSE represents the mean square error loss function in the neural network model; voltage constraint of key nodes is expressed in Eq. 16:

$$\begin{aligned} L_{\text{constraint},2} &= min\left(\frac{1}{N_u}\sum_{i=1}^{N_u}\left|U_{\text{key}} - U_{\text{key},min}\right|^2\right) \\ &= min\left(\text{MSE}_{U_{\text{key}},U_{\text{key},min}}\right) \end{aligned} \tag{16}$$

Where, N_u represents the number of key nodes; MSE represents the mean square error loss function in the neural network model.

Hence, the objective of the knowledge embedding model is twofold: not only to minimize the loss of the GAN but also to ensure that the operation scenarios generated by the model adheres to the physical constraints of key sections, guided by the incorporation of risk knowledge. Its objective function can be written as:

$$\begin{aligned} L_{\text{total}} &= \min_{G,L} \max_D W(D,G,L) \\ &= E_{x\sim p_{data}(x)}[log D(x)] \\ &\quad + E_{z\sim p_z(z)}[log(1 - D(G(z)))] + aL_{\text{constraint},i} \end{aligned} \tag{17}$$

Where a represents the hyperparameter.

4 Risk probability sampling

In the presence of a significant number of risk scenarios in the environment, the interaction between the DRL agent and the environment results in the accumulation of a substantial amount of experience in the buffer. Effectively utilizing this experience to update the intelligence becomes the second challenge addressed in this paper, as depicted in Figure 4.

To tackle this challenge, we enhance the Soft Actor-Critic (SAC) algorithm in DRL (Haarnoja Zhou et al., 2018; Christodoulou,

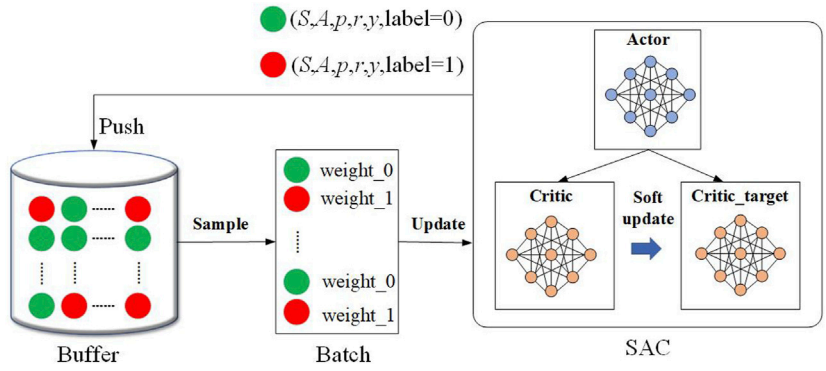

FIGURE 4
The framework of risk probability sampling.

2019). Specifically, when risk scenarios are sampled, they are marked in the buffer. During the sampling process, weights are assigned to these risky scenarios, influencing the update process of the strategy network and value network. By assigning appropriate weights, the agent can be updated more effectively, leveraging the experience gained from risky scenarios.

These improvements aim to optimize the utilization of experience stored in the buffer, allowing the DRL agent to learn from and adapt to risk scenarios, ultimately enhancing its performance in handling risk scenarios in power systems.

The traditional SAC algorithm relies on an averaging approach during the updating process, which can overlook risk scenarios stored in the buffer. This limitation hinders the agent's ability to effectively adapt to reward changes in these scenarios. To address this issue, we propose an enhancement in this paper by introducing labels to identify the risk scenarios encountered by the agent. These labels are used to assign significant weights during the network parameter update of the SAC algorithm. The specific weights are determined to meet the following constraints:

$$\begin{aligned} n_0 + n_1 &= S_{\text{batch}} \\ n_0 W_0 + n_1 W_1 &= 1 \end{aligned} \tag{18}$$

Where S_{batch} represents the size of the batch, n_0 represents the number of normal scenarios in the batch, n_1 represents the number of risk scenarios in the batch, and W_0 and W_1 represent the weights of normal and risk scenarios, respectively.

5 Case study

5.1 The construction of environment

The study in this paper enhances the existing IEEE30-node system by incorporating two wind power stations and two photovoltaic power stations. The power upper limit of the critical branch is set at 100 MW. The specific topology is illustrated in Figure 5.

In this paper, the environment comprising 4,800 operation scenarios spanning a duration of 200 days is generated using Monte Carlo simulation (Rubinstein and Kroese, 2016).

Each time-step agent achieved a maximum reward of 2, which consisted of two components: 1 reward for ensuring grid operational safety and 1 reward for optimizing grid economics. The specific reward value is set as shown in the following equations:

$$r_1 = \begin{cases} 1 & p_{\text{line}} < p_{\text{line, max}} \\ 1 - 10\dfrac{p_{\text{line}}}{p_{\text{line, max}}} & p_{\text{line, max}} \le p_{\text{line}} \le 1.1\, p_{\text{line, max}} \\ -2 & p_{\text{line}} > 1.1\, p_{\text{line, max}} \end{cases} \tag{19}$$

$$r_2 = \frac{\sum_{j=1}^{N_{\text{new}}} P_j}{\sum_{j=1}^{N_{\text{new}}} \overline{P_j}} - \frac{\sum_{i=1}^{N_{\text{Gen}}} \left(aP_i^2 + bP_i + c\right)}{\sum_{i=1}^{N_{\text{Gen}}} \left(aP_{i,max}^2 + bP_{i,max} + c\right)} \tag{20}$$

$$R = w_1 r_1 + w_2 r_2 \tag{21}$$

Where p_{line} represents the transmission power of the key line and $p_{\text{line,max}}$ represents the max transmission power of the key line. N_{Gen} and N_{new} represent the number of thermal and RES units.

The first component, grid operational safety, accounted for 1 reward point. This reward was earned by making decisions that maintained the safety and stability of the grid. If a −2 reward is earned, it will trigger an automatic termination of your policy within the grid, rendering it impossible to earn any subsequent reward.

The second component, grid economics, also contributed 1 reward point. This reward was obtained by making decisions that effectively managed and optimized the economic aspects of the grid such as promoting the rate of RES consumption.

5.2 The details of experiments

The proposed approach in this paper is subjected to three experiments to evaluate its effectiveness. Here are the details of each experiment.

1) Performance of different GAN models:

This experiment aims to assess the capabilities of KEG_GAN in generating critical scenarios while maintaining the invariance of the data distribution.

2) Different training data on the performance of KEG_GAN:

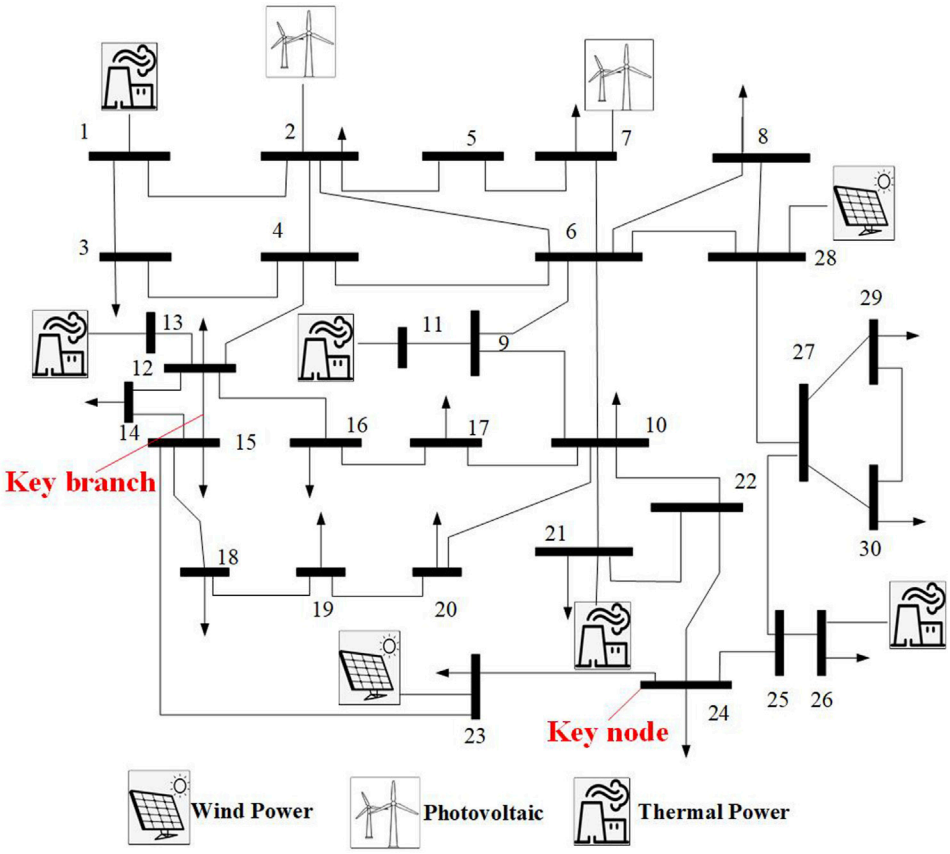

FIGURE 5
IEEE30 node system environment.

In this experiment, the impact of different training data on the performance of KEG_GAN is investigated. Datasets with different critical scenarios proportions may be used as training data, and the performance of KEG_GAN is measured and compared across these different datasets. This experiment helps determine the robustness and adaptability of KEG_GAN to different training data sources.

3) The influence of different scenarios on DRL:

The last experiment aims to assess and compare the performance of conventional DRL methods with our proposed improved DRL. By conducting a comparative analysis, valuable insights can be gained regarding the applicability and effectiveness of KEG_GAN in enhancing the performance of DRL.

5.3 Case analysis

5.3.1 The model structure of KEG_GAN

In this case, the KEG_GAN model extends the traditional GAN architecture by incorporating two additional layers of graph convolution. The generator component of the model takes a randomly sampled vector from a 200-dimensional standard normal distribution as input. It then passes through two layers of graph convolution to extract node information. Finally, a 30 × 3 matrix is outputted through a multilayer perceptron. The discriminator component of the model takes the 30 × 3 matrix as input and processes it through two layers of graph convolution and a multilayer perceptron. The final output is a scalar value representing the discriminant result. The ReLU function is employed as the activation function between the neural networks of each layer. The KEG_GAN model employs the generative adversarial loss as the loss function for the discriminator and L_{total} for the generator. It utilizes the Adam optimization algorithm to perform gradient descent and update the model parameters. Table 1 provides an overview of the model parameters.

5.3.2 Experiment 1

To assess the performance of KEG_GAN and analyze the disparity in data distribution between the generated samples and the training samples, this paper employs the KL divergence as a metric. The calculation formula of KL divergence is shown in Formula (22):

$$D_{KL}(H\|K) = \sum_{i=1}^{M}\left[h(x_i)\log h(x_i) - h(x_i)\log k(x_i)\right] \quad (22)$$

Where H is the data distribution of the guided samples, and K is the data distribution of the guiding samples. In this paper, H represents the operation scene distribution generated by the generator, and K represents the operation scene training set

TABLE 1 The model structure of KEG_GAN.

Model	Network structure	Input dimension	Output dimension
Generator	GCN	30*3 and 30*30	30*16
	GCN	30*16	30*8
	MLP	30*8	30*3
Discriminator	GCN	30*3 and 30*30	30*16
	GCN	30*16	30*8
	MLP	30*8	1

TABLE 2 KL divergence of different models.

Network	$D_{KL}(P)$	$D_{KL}(Q)$	$D_{KL}(V)$
GAN	0.2592	0.1542	0.0026
KE-GAN	0.2788	0.1513	0.0076
G-GAN	0.2087	0.1526	0.0031
KEG_GAN	0.2252	0.1596	0.0028

TABLE 3 The proportion of critical scenarios.

Network	Proportion (%)
Train data	1.37
GAN	1.54
KE-GAN	3.10
G-GAN	1.38
KEG_GAN	5.15

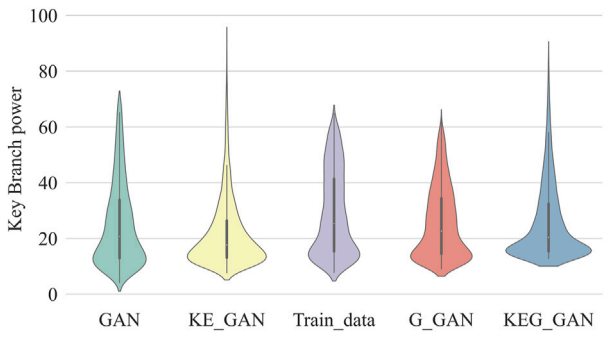

FIGURE 6
Key branch power distribution of test model.

distribution. Where the smaller KL divergence proves that the samples generated by the model are closer to the real samples.

In contrast, KEG_GAN is designed to address this issue by learning and capturing more diverse data distributions. This enables the model to generate samples that better conform to the distribution of the training data, even in scenarios characterized by higher levels of randomness and uncertainty. By leveraging the capabilities of KEG_GAN, the generated samples exhibit greater variability and better match the diversity present in the training data distribution. This allows for more accurate representation and generation of operation scenarios, particularly in scenarios with increased complexity and uncertainty introduced by the integration of RES.

As shown on Table 2, the results of our study demonstrate the enhanced sample generation capability of KEG_GAN. The incorporation of GCN enhances the information extraction capability of the model, minimizing the distance between the generated operation scenarios and real operation scenarios. Furthermore, knowledge embedding in KE-GAN leads to an increase in the KL divergence of the model. This knowledge embedding step changes the distribution of the generated data.

In the distributed network system with RES, the operation scenarios exhibit greater diversity, leading to a more varied data distribution. GAN's performance is significantly reduced in such scenarios, with the KL divergence of active power sharply increasing. This divergence indicates a significant deviation from the real data distribution, making it challenging for the generated samples to meet the requirements of intelligent algorithms.

However, the proposed KEG_GAN method in this paper addresses these challenges by representing the grid operation scenarios graphically and embedding the neural network model within the physical mechanisms. By maintaining the same data distribution, KEG_GAN achieves the generation of high-quality grid operation scenario samples.

In large-scale power grid operation scenarios, it is common to encounter a significant imbalance in sample distribution, where there are more samples representing normal operation conditions and fewer samples representing risky operation scenarios. In this paper, we address this issue by incorporating knowledge embedding into the neural network model, allowing the generated scenarios to consider the inherent risks in power grid operations. To assess the effectiveness of knowledge embedding in generating extreme operation scenarios, we focus on the power distribution of key branches as an example. To analyze the generated samples, we perform power flow calculations and examine the power distribution of these key branches. Figure 6 illustrates the power distribution of the key branches.

In particular, we set the maximum allowable transmitting power of the key branch to 100 MW. Any scenario in which the difference between the transmission power of the key branch and the maximum allowable transmission power exceeds 25 MW is

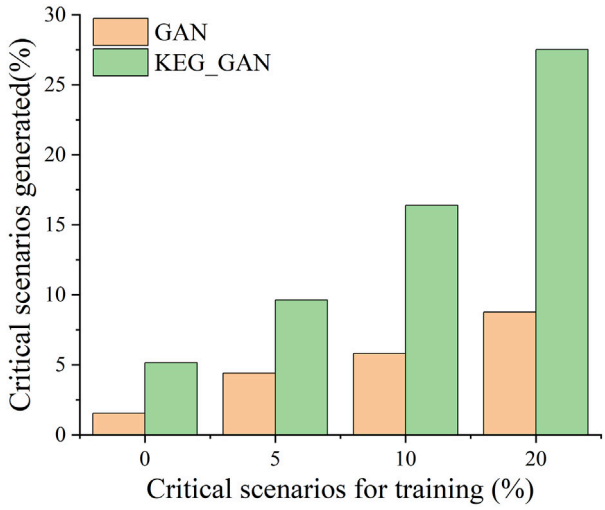

FIGURE 7
The proportion of critical scenarios generated by GAN and KEG_GAN.

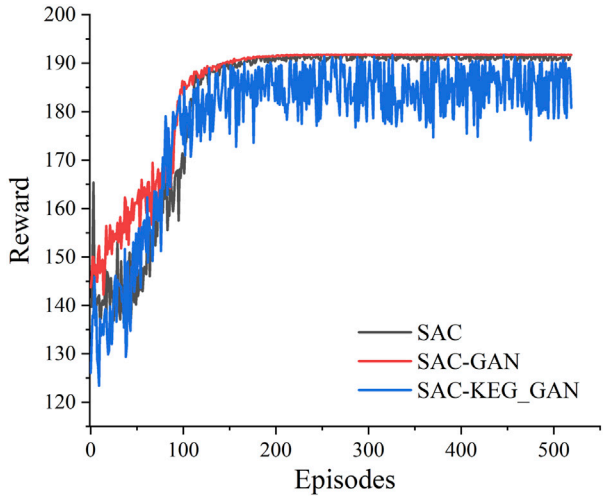

FIGURE 8
The training reward for different agents.

considered an extreme operation scenario. By evaluating the power distribution of key branches, we can gauge the capability of knowledge embedding in generating extreme operation scenarios. This analysis provides insights into the effectiveness of our approach in capturing and representing the risks associated with power grid operations.

As depicted in Figure 6, the width of the violin plot illustrates the proportion of different data. Notably, the key branch power generated by G-GAN closely aligns with the distribution of the training data. While G-GAN minimizes KL divergence, it does not yield an improvement in the performance of critical scenarios. On the other hand, KE-GAN enhances the proportion of critical scenarios in the operation scenarios. Although KE-GAN improves the performance of critical scenarios, it leads to a reduction in the performance of data distribution.

As shown on Table 3, the results indicate that KEG_GAN achieves the greatest improvement in the performance of critical scenarios while minimizing the decline in the performance of data distribution. By combining the strengths of G-GAN and KE-GAN, KEG_GAN effectively increases the proportion of extreme operation scenarios in the generated scenarios while preserving the same data distribution. This addresses the challenge of extremely sparse samples in extreme operation scenarios. KEG_GAN achieves this by incorporating basic physical constraints such as power flow and section constraints, which regulate the generated samples according to the power flow section constraint, bringing them closer to the extreme operation scenarios. Consequently, KEG_GAN offers a solution for the highly imbalanced distribution of extensive power grid operation scenarios.

5.3.3 Experiment 2

In this paper, we want to explore the effect of different training samples on the model performance so that the extreme scenario percentages of 0, 5, 10 and 20 are set to verify the extreme scenario sample generation capability of the proposed method. The proportion of critical scenarios in the generated sample is shown in Figure 7.

As the proportion of critical scenarios in the operation scenarios increases, the proportion of critical scenarios in the scenarios generated by GAN and KEG_GAN also increases. However, the performance improvement of GAN in generating critical scenarios is not significant. Additionally, the proportion of critical scenarios generated by GAN is consistently lower than the proportion of critical scenarios in the training scenarios.

On the other hand, when the proportion of critical scenarios in the training scenarios varies, KEG_GAN demonstrates an improvement in generating critical scenarios compared to the training scenarios. By leveraging knowledge embedding to incorporate features of critical scenarios, KEG_GAN directs the agent's focus towards critical scenarios, thereby increasing their proportion in the generated scenarios. In contrast, GAN tends to prioritize the data distribution in the operation scenarios and tends to neglect the extreme operation scenarios, resulting in a decreasing proportion of extreme operation scenarios in the generated scenarios. Consequently, relying solely on increasing the proportion of critical scenarios for GAN to generate critical scenarios often proves ineffective.

Furthermore, when there are no critical scenarios present in the training samples, KEG_GAN exhibits the capability to generate approximately 5% of critical scenarios. This demonstrates that our method possesses few-shot (Sung et al., 2018) or zero-shot (Xian et al., 2018) capabilities, while GAN struggles to generate unseen samples.

5.3.4 Experiment 3

The training data comprised different sets of samples. Following the completion of training, the agents were tested over a period of 10 consecutive days, with decision-making intervals of 15 min for both training and testing phases. The training result is presented in Figure 8.

During the training process, we observed that the reward of the agent trained solely on real data was not significantly different from the agent trained using data generated by GAN. However, SAC trained with real data contained a small number of infrequent

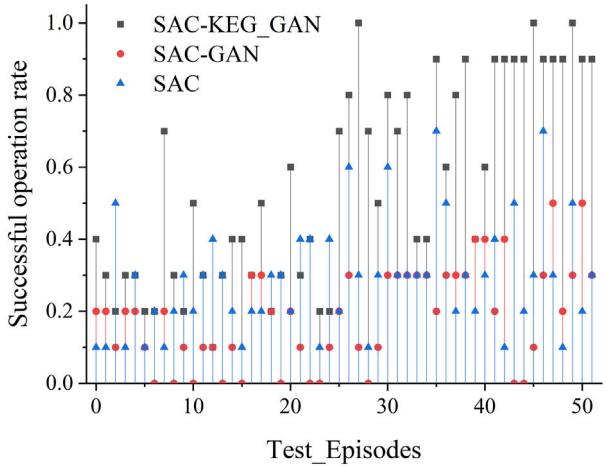

FIGURE 9
The successful operation rate for different agents.

TABLE 4 The Reward of different agents.

Day	SAC	SAC-GAN	SAC-KEG_GAN
1	189.5729	180.5112	175.8437
2	158.3673	179.2283	179.1771
3	187.39	178.6056	177.0162
4	187.6509	179.9945	176.9828
5	188.0185	130.5518	171.3925
6	128.6596	118.0459	172.5115
7	164.0986	32.2041	171.3697
8	184.9219	159.9455	174.9195
9	187.6598	50.10933	174.6427
10	188.1027	188.3726	153.0167

critical scenarios, which were sampled less frequently. As a result, the overall training process exhibited minimal fluctuations. On the other hand, SAC-GAN tended to overlook such critical scenarios, leading to smoother loss curves for the agent. Unfortunately, this smoothness also made it difficult for the agent to adequately account for these critical scenarios.

By incorporating KEG_GAN enhanced data into the training process, we enable the SAC-KEG_GAN to explore a broader range of risk scenarios. As a result, the training reward exhibits oscillations when compared to the traditional SAC algorithm.

This oscillation is challenging to achieve when relying solely on raw data. Consequently, SAC focuses on minimizing the training cost, thereby attaining a stable reward.

To evaluate the performance of various agents, we carried out a comprehensive 10-day testing phase. During this period, the agents actively responded to the changing environmental conditions by making decisions every 15 min. In a single day, there were a total of 96 time sections in which decisions were made. One crucial aspect we assessed was the impact of key branch crossings on grid safety. If an agent's decision resulted in crossing the safety limit of the grid, it rendered the grid unsafe, and the agent was unable to continue participating in the decision-making process. The test results, presented in Figure 9 and Table 4, provide a clear visualization of the agents' performance throughout the testing phase.

It becomes evident that as the number of training iterations increases, both SAC and SAC-GAN fail to ensure the safe and stable operation of the grid. The test consistently gets interrupted on certain days due to the key branch surpassing its limit. Consequently, the smart body is unable to receive subsequent rewards. However, SAC-KEG_GAN incorporates a comprehensive consideration of the risk associated with grid operations. It evaluates both the risk and the economic aspects of the grid, enabling it to provide a more robust strategy. After a certain number of training iterations, the decisions made by SAC-KEG_GAN lead to a grid that can operate safely and steadily for a duration of 10 days. In the analysis, the best strategies from the aforementioned testing process were selected and their results are presented in Table 4. The best successful operation rate of SAC is 0.7, SAC-GAN achieves a successful operation rate of 0.5, while SAC-KEG_GAN demonstrates a successful operation rate of 1.

Upon closer examination, it is discovered that SAC fails to obtain subsequent rewards on days 2, 6, and 7 due to the given decision's critical section crossing its limit. On the contrary, the performance of SAC-GAN is marginally inferior to that of SAC. This can be attributed to SAC-GAN's tendency to overlook samples from high-risk scenarios during the process of data augmentation. As a consequence, the generated scenarios might lack the critical instances that contribute to the overall performance of the policy learned by the intelligent agent. Although the cumulative reward of the strategy provided by SAC-KEG_GAN may be lower than that of SAC on certain days, it effectively evaluates the risk and economic aspects of grid operation by learning from experiences gained in risky scenarios. As a result, it generates decisions that enable the grid to operate safely and improve the economics of grid operation while ensuring grid safety.

6 Conclusion

In the context of high-dimensional uncertainty, this paper addresses the limited adaptability of policies in critical risk scenarios. By adopting a multi-objective modeling approach that incorporates both security and economy, the original problem is formulated as a multi-stage risk-averse stochastic sequential decision-making problem with dynamic risk metrics. To tackle this challenge, Risk-averse stochastic dynamic dispatch of power systems based on deep reinforcement learning with risk-oriented Graph-Gan sampling is proposed. This policy aims to overcome the shortcomings of existing methods in risk sample generation and scenarios identification, enabling the rapid solution of optimal risk-averse intraday dispatch policy. Simulation results demonstrate that proposed approach outperforms other commonly used online dispatch policies, which not only improves the economic efficiency of power system operations

but also reduces the potential high costs associated with critical scenarios. This algorithm incorporates risk-averse preferences to avoid unnecessary load shedding, particularly in scenarios involving RESs abandonment. Hence, it is crucial to carefully consider the risk aversion preferences of the algorithm in the specific application. Furthermore, the proposed algorithm achieves high computing efficiency in real-time scheduling through offline learning and does not rely on predictive information for real-time scheduling. Its promising application prospects and scalability extend to addressing other complex online stochastic optimization scheduling problems in future smart grids.

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

WH: Writing–original draft. ZD: Writing–original draft. JH: Writing–original draft. LL: Writing–original draft. YC: Writing–original draft. ZC: Writing–review and editing. ZP: Writing–original draft.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. The authors declare that this study received funding from the science and technology project of China Southern Power Grid (Project number: 670000KK52210021).

Conflict of interest

Authors WH, ZD, JH, LL, and YC were employed by China Southern Power Grid. Guangzhou

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.

References

Arjovsky, M., and Bottou, L. (2017). "Towards principled methods for training generative adversarial networks," in Proceedings of the 5th International Conference on Learning Representations, Toulon, France, April 2017.

Baran, M. E., and Wu, F. (1989). Network reconfiguration in distribution systems for loss reduction and load balancing. *IEEE Trans. Power Deliv.* 4 (2), 1401–1407. doi:10.1109/61.25627

Bedoya, J. C., Wang, Y., and Liu, C.-C. (2021). Distribution system resilience under asynchronous information using deep reinforcement learning. *IEEE Trans. Power Syst.* 36 (5), 4235–4245. doi:10.1109/tpwrs.2021.3056543

Chen, J., Yu, T., Yin, L., Tang, J., and Wang, H. (2019). A unified time scale intelligent control algorithm for micro grid based on extreme dynamic programming. *CSEE J. Power Energy Syst.* (99), 1–7. doi:10.17775/CSEEJPES.2019.00100

Chen, M., Shen, Z., Wang, L., and Zhang, G. (2022). Intelligent energy scheduling in renewable integrated microgrid with bidirectional electricity-to-hydrogen conversion. *IEEE Trans. Netw. Sci. Eng.* 9 (4), 2212–2223. doi:10.1109/tnse.2022.3158988

Chen, Y., Wang, Y., Kirschen, D., and Zhang, B. (2018). Model-free renewable scenario generation using generative adversarial networks. *IEEE Trans. Power Syst.* 33 (3), 3265–3275. doi:10.1109/tpwrs.2018.2794541

Christodoulou, P. (2019). Soft actor-critic for discrete action settings. Available at: https://arxiv.org/abs/1910.07207.

Goodfellow, I. J., Pouget-abadie, J., Mirza, M., Xu, B., Warde-Farley, D., and Ozair, H. (2014). "Generative adversarial nets," in Proceedings of the 27th International Conference on Neural Information Processing Systems, Cambridge, MA, December 2014 (MIT Press), 2672–2680.

Guan, J., Tang, H., Ke, W., Yao, J., and Yang, S. (2020). A parallel multi-scenario learning method for near-real-time power dispatch optimization. *Energy* 202, 117708. doi:10.1016/j.energy.2020.117708

Guo, C., Wang, X., Zheng, Y., and Zhang, F. (2022). Real-time optimal energy management of microgrid with uncertainties based on deep reinforcement learning. *Energy* 238, 121873–121885. doi:10.1016/j.energy.2021.121873

Haarnoja, T., Zhou, A., Hartikainen, K., Tucker, G., Ha, S., and Tan, J. (2018). Soft actor-critic algorithms and applications. Available at: https://arxiv.org/abs/1812.05905.

Hua, H., Qin, Y., Hao, C., and Cao, J. (2019). Optimal energy management strategies for energy Internet via deep reinforcement learning approach. *Appl. Energy* 239, 598–609. doi:10.1016/j.apenergy.2019.01.145

Liu, W., Zhuang, P., Liang, H., and Peng, J. (2018). Distributed economic dispatch in microgrids based on cooperative reinforcement learning. *IEEE Trans. Neural Netw. Learn. Syst.* 29 (6), 2192–2203. doi:10.1109/tnnls.2018.2801880

Lv, K., Tang, H., Bak-Jensen, B., Radhakrishna Pillai, J., Tan, Q., and Zhang, Q. (2020). Hierarchical learning optimisation method for the coordination dispatch of the inter-regional power grid considering the quality of service index. *Transm. Distribution* 14 (18), 3673–3684. doi:10.1049/iet-gtd.2019.1869

Mathiesen, B., Lund, H., Connolly, D., Wenzel, H., Østergaard, P., Möller, B., et al. (2015). Smart Energy Systems for coherent 100% renewable energy and transport solutions energy and transport solutions. *Appl. Energy* 145, 139–154. doi:10.1016/j.apenergy.2015.01.075

Pan, Z., Yu, T., Li, J., Qu, K., and Yang, B. (2020). Risk-averse real-time dispatch of integrated electricity and heat system using a modified approximate dynamic programming approach. *Energy* 198, 117347. doi:10.1016/j.energy.2020.117347

Rubinstein, R. Y., and Kroese, D. P. (2016). *Simulation and the Monte Carlo method.* Hoboken, NJ, USA: John Wiley and Sons.

Shervashidze, N., Vishwanathan, S. V. N., Petri, T. H., Mehlhorn, K., and Borgwardt, K. M. (2009). "Efficient graphlet kernels for large graph comparison," in *Proc. Of the 12th int'l conf. On artificial intelligence and statistics* (Clearwater: MIT Press), 488–495.

Silver, D., Huang, A., Maddison, C. J., Guez, A., Sifre, L., Van Den Driessche, G., et al. (2016). Mastering the game of Go with deep neural networks and tree search. *Nature* 529 (7587), 484–489. doi:10.1038/nature16961

Sung, F., Yang, Y., Zhang, L., Xiang, T., Torr, P. H., and Hospedales, T. M. (2018). "Learning to compare: relation network for few-shot learning," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, June 2018, 1199–1208.

Thomas, N. K., and Max, W. (2016). Semi-supervised classification with graph convolutional networks. Available at: https://arxiv.org/abs/1609.02907.

Xian, Y., Lorenz, T., Schiele, B., and Akata, Z. (2018). "Feature generating networks for zero-shot learning," in Proceedings of the IEEE conference on computer vision and pattern recognition, Salt Lake City, UT, USA, June 2018, 5542–5551.

Yu, T., Wang, H., Zhou, B., Chan, K. W., and Tang, J. (2015). Multi-agent correlated equilibrium Q(λ) learning for coordinated smart generation control of interconnected power grids. *IEEE Trans. Power Syst.* 30 (4), 1669–1679. doi:10.1109/tpwrs.2014.2357079

Zhang, C., Shao, Z., Jiang, C., and Chen, F. (2021b). A PV generation data reconstruction method based on improved super-resolution generative adversarial network. *Int. J. Electr. Power and Energy Syst.* 132, 107129. doi:10.1016/j.ijepes.2021.107129

Zhang, M., Chen, J., Yang, Z., Peng, K., Zhao, Y., and Zhang, X. (2021a). Stochastic day-ahead scheduling of irrigation system integrated agricultural microgrid with pumped storage and uncertain wind power. *Energy* 237, 121638. doi:10.1016/j.energy.2021.121638

Zhao, T., and Wang, J. (2021). Learning sequential distribution system restoration via graph-reinforcement learning. *IEEE Trans. Power Syst.* 37 (2), 1601–1611. doi:10.1109/tpwrs.2021.3102870