INDERSCIENCE PUBLISHERS
Linking academia, business and industry through research

International Journal of Critical Infrastructures

ISSN online: 1741-8038 - ISSN print: 1475-3219
https://www.inderscience.com/ijcis

Knowledge graph and reinforcement learning-based automatic decision-making for distribution network fault handling

Wenshuo Lin, Jiawei Ai, Weixiang Zhang, Yiming Peng, Huili Tian, Chen Wang

DOI: 10.1504/IJCIS.2026.10081439

Article History:

Received:	13 July 2026
Last revised:	27 August 2026
Accepted:	28 August 2026
Published online:	23 October 2026

Copyright © 2026 Inderscience Enterprises Ltd.

Knowledge graph and reinforcement learning-based automatic decision-making for distribution network fault handling

Wenshuo Lin*, Jiawei Ai, Weixiang Zhang, Yiming Peng, Huili Tian and Chen Wang

Guangzhou Power Supply Bureau,
Guangdong Power Grid Co., Ltd.,
Guangzhou, 510620, China
Email: linwenshuo@gzps.corp.csg
Email: aijiawei0126@gzps.corp.csg
Email: zhangweixiang1212@gzps.corp.csg
Email: pengyiming@gzps.corp.csg
Email: tianhuili@guangzhou.csg.cn
Email: wangchen1226@gzps.corp.csg
*Corresponding author

Abstract: To address the low decision accuracy, long decision time and limited fault recovery rate of traditional distribution network fault decision methods, this paper proposes an automatic decision-making method for distribution network fault handling based on knowledge graph and reinforcement learning. Firstly, the canopy clustering algorithm and UMAP manifold learning are adopted to collect and preprocess operational data of distribution networks. A top-down fault handling knowledge graph with pattern and data layers is further constructed to realise dynamic knowledge updating. Moreover, a knowledge graph-embedded deep Q-network decision framework is designed, which generates optimal fault handling sequences through offline dual-strategy learning. Experimental results verify that the proposed method achieves stable decision accuracy of 97.8–98.9% within 0.52–0.71 s, with a maximum fault recovery rate of 99.12%.

Keywords: distribution network fault handling; knowledge graph; reinforcement learning; canopy clustering; deep Q-network.

Reference to this paper should be made as follows: Lin, W., Ai, J., Zhang, W., Peng, Y., Tian, H. and Wang, C. (2026) 'Knowledge graph and reinforcement learning-based automatic decision-making for distribution network fault handling', *Int. J. Critical Infrastructures*, Vol. 22, No. 14, pp.1–24.

Biographical notes: Wenshuo Lin graduated from Shanghai Jiao Tong University,majoring in Electrical Engineering and Automation. He currently works as an Engineer at the Power Dispatching Control Center of Guangzhou Power Supply Bureau. His main research direction is distribution network dispatching operation and control.

Jiawei Ai obtained his Master's degree in Engineering from South China University of Technology in Guangzhou, China in 2022. He is an Engineer of China Southern Power Grid Group Co., Ltd. in Guangdong Province, China. His main research direction is power grid scheduling and operation.

Copyright © The Author(s) 2026. Published by Inderscience Publishers Ltd. This is an Open Access Article distributed under the CC BY license. (http://creativecommons.org/licenses/by/4.0/)

Weixiang Zhang obtained his Master's degree in Engineering from North China Electric Power University in Beijing, China in 2024. He is an Engineer of China Southern Power Grid Group Co., Ltd. in Guangdong Province, China. His main research direction is power grid scheduling and operation.

Yiming Peng holds a Bachelor's degree in Electrical Engineering and Automation from the School of Electric Power at South China University of Technology. Currently, she is a technical staff member of the Guangzhou Power Supply Bureau Control Center of Guangdong Power Grid Co., Ltd., mainly engaged in the construction of new systems and functions for new power systems, as well as technological innovation, such as the automation and informatisation of main distribution network dispatching.

Huili Tian received her Master's degree and is currently a Senior Engineer. Her research interests include distribution network dispatching and operation, resilient power grids, and disaster prevention and mitigation technologies, with a focus on technology implementation and demonstration projects.

Chen Wang holds a Master's degree from the School of Electrical Engineering at Xi'an Jiaotong University. She currently works at the Power Dispatching and Control Center of Guangzhou Power Supply Bureau, Guangdong Power Grid, with research focusing on distribution network control technology.

1 Introduction

With the accelerated construction of new power systems, distributed power sources, energy storage devices, and various diversified loads have achieved large-scale integration into the distribution network (Wang and Sun, 2025). In this context, the amplitude and direction of fault currents become complex and difficult to predict, and the fault evolution process becomes more rapid, with significant randomness in its impact range (Zhou et al., 2026). At the same time, users' requirements for power supply reliability and power quality continue to increase, and the frequent occurrence of extreme weather events further exacerbates the frequency and difficulty of handling distribution network failures. In this context, the handling of distribution network faults faces practical challenges such as large information processing volume, strong decision-making timeliness, and variable operation modes. It is urgent to improve the automation and intelligence level of post fault isolation, recovery, and reconstruction (He et al., 2025). In this context, the research on automatic decision-making for fault handling in distribution networks aims to use real-time measurement information and intelligent optimisation algorithms to achieve rapid identification of fault types, accurate positioning of fault sections, automatic generation of power supply recovery paths, and coordinated control between distributed power sources and the main grid. This has important theoretical value and engineering practical significance for ensuring the safe and stable operation of distribution networks and promoting the automation level of smart grids (Bera and Mande, 2025; Li et al., 2025b).

In response to the above research topic, Zhang et al. (2024) proposed a distribution network fault recovery decision-making method based on graph reinforcement learning. Using graph data to fully depict the key decision-making elements in the fault recovery process, covering the topology connection relationship and electrical quantity

characteristics of the distribution network. Design a graph reinforcement learning architecture, with a pre graph neural network module responsible for processing dynamically changing graph data inputs, effectively adapting to frequent topology adjustments in the fault recovery phase of the distribution network. The graph neural network built into the reinforcement learning agent directly generates the optimal fault recovery plan, significantly improving decision response speed. However, graph reinforcement learning models have insufficient generalisation ability for small sample fault scenarios, with reward functions biased towards decision speed rather than accuracy, and insufficient integration of multi-source information such as protection actions and fault recording, resulting in a decrease in the accuracy of automatic decision-making for distribution network fault handling under complex fault conditions. Song et al. (2025) proposed a distribution network fault recovery decision method based on BERT-BiGRU-CRF. A deep learning model that integrates BERT, BiGRU and CRF is designed specifically to extract key knowledge from fault texts in distribution networks, addressing the problem of poor entity recognition performance due to small sample sizes. Secondly, in order to integrate power data with diverse sources and inconsistent structures, a knowledge graph for fault dispatch and disposal in distribution networks was constructed. Heterogeneous information was uniformly converted into structured knowledge forms, and the visualisation of the graph was achieved using the Neo4j graph database. Finally, it was applied to the dispatch assisted decision-making process for fault disposal in distribution networks. The model structure is complex and computationally intensive, coupled with additional delays in the Neo4j retrieval and inference process of the knowledge graph, making it difficult for the overall serial processing flow to meet real-time response requirements. Xia et al. (2024) proposed a distribution network fault recovery decision method based on mixed integer programming. Regarding the optimisation and arrangement of long-term maintenance plans for distribution network feeders, fault risks are divided into three categories, and corresponding operation and maintenance measures are matched for each category of risk. On this basis, a multi-objective mixed integer linear programming model is established to achieve high-precision lean scheduling of distribution network maintenance plans. After model solving and optimisation, the economic indicators of operation and maintenance as well as the reliability indicators of power supply in the distribution system have been effectively improved. However, the mixed integer linear programming model used is oriented towards long-term maintenance scheduling rather than real-time fault recovery, simplifying the topology structure and omitting some transfer switch combinations. At the same time, conservative optimisation strategies tend to abandon edge loads to reduce computational complexity, resulting in a decrease in the fault recovery rate of its distribution network.

A new automatic decision-making method for distribution network fault handling based on knowledge graph and reinforcement learning is proposed with the expected goal of solving the problems of low accuracy, long automatic decision-making time, and low recovery rate of distribution network fault handling in the above methods. This route first uses canopy clustering algorithm to quickly screen multi-source heterogeneous operation data of distribution network, and then adopts UMAP manifold learning method to preserve the global topology structure of high-dimensional data in the dimensionality reduction process, in order to solve the scalability and structural fidelity problems in massive real-time data; subsequently, a top-down two-level knowledge graph is

constructed, using the BERT BiLSTM CRF model to extract entities and relationships from unstructured and semi-structured scheduling rules, fault handling records, and defect libraries, forming a hierarchical architecture that encapsulates rule constraints at the pattern layer and stores specific instances at the data layer. A separation mechanism is designed for triggering pattern layer updates and periodically triggering data layer updates to ensure the timeliness of knowledge; finally, a deep Q-network decision framework that integrates knowledge graphs is designed. Topological semantics and measurement data are used as state inputs, and regulatory constraints are incorporated into the reward function. The optimal fault handling action sequence is automatically generated using the offline learning strategy of the target network and valuation network. The core feature of this method is the construction of a complete closed loop that integrates data preprocessing, knowledge modelling, and reinforcement learning decision-making. Its novelty is reflected in three aspects: firstly, the introduction of UMAP's Riemannian geometry and algebraic topology theory in data dimensionality reduction to solve the problem of uneven distribution of distribution network operation data during peak and low load periods, which is different from traditional linear dimensionality reduction methods such as principal component analysis; secondly, the dual level update mechanism of knowledge graph enables structured prior knowledge to continuously evolve with the revision of regulations and the accumulation of fault cases, which is different from the static construction mode of conventional knowledge graph; thirdly, in deep Q-networks, it is not simply a matter of applying a general network structure. Instead, knowledge graphs are embedded into the state space and reward functions are customised for distribution network fault handling scenarios, enabling reinforcement learning to maintain high decision accuracy even under small sample and topology changing conditions, thus achieving performance breakthroughs in accuracy, timeliness, and power recovery capabilities simultaneously.

2 Design of automatic decision-making method for distribution network fault handling

2.1 Data collection and preprocessing for distribution network operation

The canopy clustering algorithm is a fast and rough preprocessing clustering method. Distribution network data has the characteristics of high dimensionality, massive volume, strong real-time performance, and the presence of a large number of outliers and missing values (Al-Jumaili et al., 2024). The canopy algorithm can be used for fast data screening in the front-end collection, such as performing canopy coarse clustering on time-series data such as voltage, current, and power in distributed collection terminals. Due to the overlap between clusters, the algorithm can adapt to the fuzzy boundary characteristics of multi node power supply or load fluctuations in the distribution network, and has good tolerance for outliers in the collected data. However, it should be noted that the two thresholds of this algorithm rely on manual experience settings. If the dynamic range of distribution network data changes dramatically, it may lead to overly rough cluster partitioning or loss of effective information. Therefore, it is more suitable as a lightweight leading step in the distribution network data preprocessing chain, combined with more accurate clustering or classification methods to complete the final analysis.

Canopy clustering is a method of roughly partitioning data through central points. This method takes each canopy centre as the centre and draws small and large circles with radii of T_2 and T_1, respectively. The data points falling within the small circle range are included in the current canopy; points located outside the small circle and inside the large circle are considered potential new canopy centres and can continue to be used for subsequent data partitioning (Zhang, 2024). The specific execution process of this algorithm is as follows:

1 Set the distribution network operation dataset List, determine thresholds T_1 and T_2, and ensure that $T_1 > T_2$.

2 Randomly select a data point and mark it as A, then delete it from the data list List. Based on point A, establish an independent canopy set.

3 Calculate the distance d from all points in the set List to data point A.

4 If $T_2 < d < T_1$, then take the corresponding data point as the weak correlation point of data point A.

5 If $d < T_2$, take the data point as the strongly correlated point of data point A, and remove it from the set List, no longer selecting it as the centre point.

6 At this point, ϑ centre points will be obtained, and the clusters containing ϑ centre points will be traversed. If the number of points in a cluster is less than a certain number, it is considered that there are noise points here, so the centre point and its cluster will be directly deleted.

The collected operational data of the distribution network is shown in Table 1.

The distribution network operation data collected in this study covers three categories: structured, semi-structured and unstructured, including distribution network equipment ledger data, fault handling data, dispatching regulation data and distribution network defect database data. The data types and selected fields are listed in Table 1. The main difficulties in data acquisition are reflected in three aspects: first, multi-source heterogeneous data are scattered in different business systems such as distribution automation system, data acquisition and monitoring control system, fault recording device, protection information management subsystem and meteorological monitoring system, and the data format, sampling frequency and interface protocol of each system are not uniform, which makes cross-system association and integration difficult; second, unstructured fault handling texts and scheduling rules data contain a large number of technical terms, colloquial expressions and default information, and the entity boundaries are vague, so manual labelling is expensive and consistency is difficult to guarantee; third, the operation data of distribution network has the characteristics of high dimension, massive real-time and contains many abnormal values and missing values. Communication interruption or equipment abnormality in the field acquisition process often leads to incomplete data, which affects the quality of subsequent modelling. In view of the above difficulties, the collection method adopted has the following characteristics: relying on multiple in-service systems to realise automatic collection and ensure the authenticity and continuity of data sources; the unstructured text is marked with BIO sequence by YEDDA tool, which provides training samples for subsequent BERT-BiLSTM-CRF knowledge extraction; in data preprocessing, the principle of triple standard deviation is used to identify and correct outliers, the nearest neighbour

interpolation is used to achieve time alignment of multi-source data, the second-order low-pass Butterworth filter is used to remove high-frequency noise, and the maximum-minimum method and standard score normalisation are used to standardise electrical and non-electrical data respectively, thus laying a foundation for subsequent knowledge map construction and strengthening learning decision-making while ensuring data quality.

Table 1 Distribution network operation data

Data name	*Data type*	*Selected parts*
Distribution network equipment ledger data	Semi-structured	Including the official name of the equipment, the power station it belongs to, the operation and maintenance team, the specific person in charge, the designed voltage level, the corresponding main feeder line, and the actual date of commercial operation
Fault handling data	Unstructured	Including the name of the power outage line, fault category, cause analysis, secondary classification of fault causes, location of the power outage area, fault overview, accident handling process, and root cause explanation, etc
Dispatch protocol data	Unstructured	Including the integration of new facilities into operation management, management of power plant grid connection processes, coordination of large-scale user side grid connection, execution of dispatch regulations, system operation switching operations, emergency response mechanisms for accidents, load regulation management, maintenance plan organisation, etc
Distribution network defect database data	Semi-structured	Including equipment identification, construction team, classification of defect severity, specific description of defects, equipment category, etc

In order to solve the scalability problem caused by large-scale collection points such as smart meters and feeder terminal units in distribution network operation data, and to preserve more global topology structure in high-dimensional operation data such as voltage phase and harmonic distortion rate of each node. UMAP uses Riemannian geometry and algebraic topology to solve the problem of uneven distribution of operation data in different regions of the actual physical manifold of the distribution network, such as significant differences in data density between peak and low load periods (Li et al., 2025a). Specifically, it introduces a local connectivity constraint ρ_i to measure the distance from x_i to its nearest neighbours, and a normalisation factor σ_i to measure the total similarity from x_i to its k nearest neighbours. The calculation formula for ρ_i is as follows:

$$\rho_i = \min\left\{\left\|x_i - x_j\right\|_2 \middle| 1 \le j \le k\right\} \tag{1}$$

The calculation formula for the conditional probability $p_{j|i}$ of measuring the similarity between x_i and its neighbour x_j in high-dimensional space is as follows:

$$p_{j|i} = \exp\left(-\frac{\left\|x_i - x_j\right\|_2 - \rho_i}{\sigma_i}\right) \tag{2}$$

In the above formula, σ_i represents the local connection parameter. After calculating the conditional probability of $p_{j|i}$ for each data point, UMAP uses a symmetric matrix P to measure the distance between high-dimensional data points. The calculation formula for each element p_{ij} in matrix P is as follows:

$$p_{ij} = p_{j|i} + p_{i|j} - p_{j|i}p_{i|j} \tag{3}$$

In the above formula, p_{ij} represents the conditional probability that data point x_i is considered as its neighbour when using data point x_j as the reference.

UMAP assumes that high-dimensional data points are uniformly distributed on a potential low dimensional manifold. The probability of y_i having y_j as its neighbour in the embedding space is as follows:

$$q_{ij} = \left(1 + a'\left\|y_i - y_j\right\|_2^{2b'}\right)^{-1} \tag{4}$$

In the above formula, $a' > 0$ and $b' > 0$ represent two different hyperparameters, which are set by default to $a' \approx 1.929$ and $b' \approx 0.7915$, respectively (Li and Chauhan, 2025). It is worth noting that when $a' = 1$ and $b' = 1$, the low dimensional embedding q_{ij} is a student t-distribution. After obtaining probabilities p_{ij} between high-dimensional data points and q_{ij} between low dimensional data points, UMAP attempts to minimise the difference between p_{ij} and q_{ij} through a cross entropy loss function to find the representation of low dimensional data:

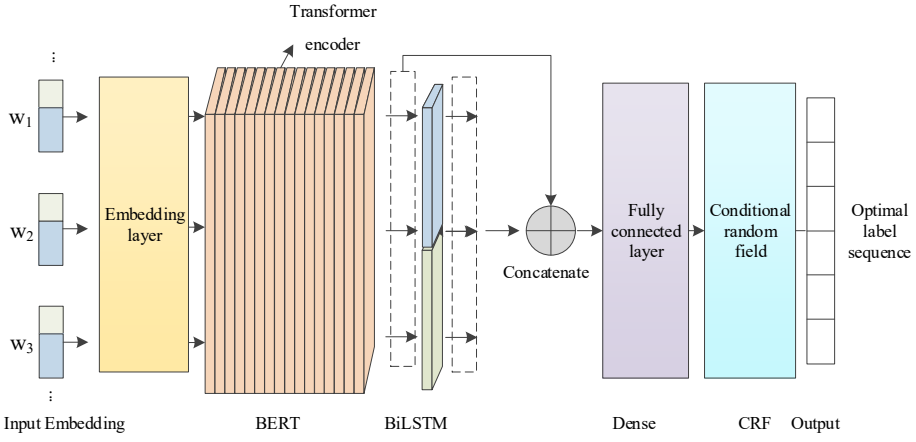
$$W = \sum_{j \neq i}\left(p_{ij}\log\left(\frac{p_{ij}}{q_{ij}}\right) + \left(1 - p_{ij}\right)\log\left(\frac{1 - p_{ij}}{1 - q_{ij}}\right)\right) \tag{5}$$

After canopy clustering and UMAP dimensionality reduction, the redundant noise and outliers are eliminated, and the high-dimensional global topological structure is kept in the low-dimensional space, which provides a clear structure and complete information input basis for subsequent knowledge extraction. However, numerical data can only reflect the state characteristics of distribution network operation, but cannot express the logical association between devices, the constraint relationship of operation procedures and the pattern law of historical faults. Therefore, it is necessary to further build a knowledge map for fault handling on the basis of data preprocessing, and store and organise these regular and empirical knowledge in the form of structured triples. Section 2.2 will detail the construction process and two-level update mechanism of knowledge map.

2.2 Construction of knowledge graph for distribution network fault handling

BERT is essentially a pre trained model for language representation. Its core architecture is built on the basis of a bidirectional Transformer and introduces an innovative masked language model (MLM) mechanism. Using this mechanism, BERT can generate deep level bidirectional semantic representations to support knowledge extraction tasks in distribution network operation data. The network structure of the knowledge extraction model for distribution network operation data is shown in Figure 1.

Figure 1 Network structure of knowledge extraction model for distribution network operation data (see online version for colours)



The network structure of the knowledge extraction model of distribution network operation data shown in Figure 1 adopts the design of series integration of BERT layer and BiLSTM layer, and its structural advantages are embodied in the following aspects. Pre-trained BERT model is loaded at the front end, and by using its bidirectional Transformer encoder architecture and mask language model pre-training mechanism, the deep semantic representation ability can be acquired on large-scale general corpus, which makes the model obtain better generalisation performance under the condition of only using small-scale manual labelling training data, effectively reducing the engineering difficulty of high cost of labelling distribution network fault text. At the same time, the BERT layer adds a [CLS] tag at the beginning and a [SEP] tag at the end of the input sequence and performs word-level segmentation, and maps the character embedding coding, position information coding and sentence information coding into the same high-dimensional space and then adds them together, so that the model can simultaneously capture multiple information of character semantics, position order and sentence boundary. BiLSTM layer is connected in series at the back end to extract the contextual linguistic features of the vector sequence output by BERT for the second time. BiLSTM's bi-directional structure can utilise both forward and backward contextual information at each time step, which makes up for the deficiency that one-way LSTM can only capture the one-way dependence of the sequence. It is especially suitable for the characteristics of strong correlation of technical terms and long context dependence of entity boundaries in distribution network fault texts, for example, multiple device names and actions can only be fully expressed in fault cause description. The model does not simply adopt serial output between BERT layer and BiLSTM layer, but designs concatenate feature series layer to splice and fuse the vector sequences generated in the two stages, so that the features that finally enter the fully connected layer include both the deep bidirectional semantic representation provided by BERT and the local temporal context features provided by BiLSTM. Compared with the single model scheme using only BERT or BiLSTM, this fusion strategy can give consideration to global semantic understanding and local sequence dependency capture, and enhance the robustness of

expression and omission of distribution network technical terms. Finally, the CRF layer is connected, which automatically constructs the state transition matrix during the learning process, and uses the transition probability to model the constraint relationship between adjacent tags, so as to ensure that the output entity tag sequence meets the reasonable transition rules such as B tag followed by I tag in the BIO tagging system, and avoid the illegal tagging sequence that may be generated when the fully connected layer independently predicts each character tag, thus improving the accuracy and reliability of knowledge extraction as a whole.

The overall framework is constructed by loading pre trained BERT models on the front-end and concatenating BiLSTM-CRF networks on the back-end. The features output by the BERT layer are concatenated and fused with the output features of the BiLSTM layer, and fine tuned training is performed on this basis. This method can effectively reduce the dependence on large-scale manually annotated training data, thereby obtaining model parameters with relatively ideal performance under the condition of using only a smaller training set (Liu et al., 2025).

1 Original input layer: use YEDDA tool to add BIO sequence labels to sentences. In this annotation system, *B* marks the first character of the entity, *I* marks the last character of the entity, and *O* represents content that does not belong to any entity.

2 BERT layer: the input text is first subjected to word level segmentation processing, and a start mark [CLS] and an end mark [SEP] are added at the beginning and end of the segmented sequence, respectively; subsequently, through the action of the embedding layer, the original word sequence $w = [w_1, w_2, \ldots, w_n]$ is mapped into a neural network analysable vectorised output sequence X. At this point, the following relationship is satisfied:

$$\begin{cases} X = [x_1, x_2, \ldots, x_n] \in R^{n \times H} \\ x_i = E_i^t + E_i^p + E_i^s, \quad i = 1, 2, \ldots, n \end{cases} \tag{6}$$

In the above equation, H represents the dimension of the vector; E_i^t represents character embedding encoding, E_i^p represents positional information encoding, and E_i^s corresponds to sentence level information encoding. After mapping these three encoding parts to a high-dimensional space of the same dimension, they are combined through an addition operation to generate input vector x_i. Next, input vector sequence X into the pre trained BERT model to extract deep semantic information from the text and generate output vector sequence $T = [t_1, t_2, \ldots, t_n]$.

In the pre training stage of BERT, the model relies on large-scale universal corpora and adopts the mask language model (MLM) mechanism for unsupervised learning to achieve strong generalisation ability. This process will randomly mask about 15% of the words in the input sequence, and then use a Transformer encoder to read the entire word sequence at once, and use the contextual information of each word to predict the masked part. At the same time, the pre training phase will also introduce next sentence prediction (NSP) tasks to determine whether there is a coherent contextual connection between two sentences.

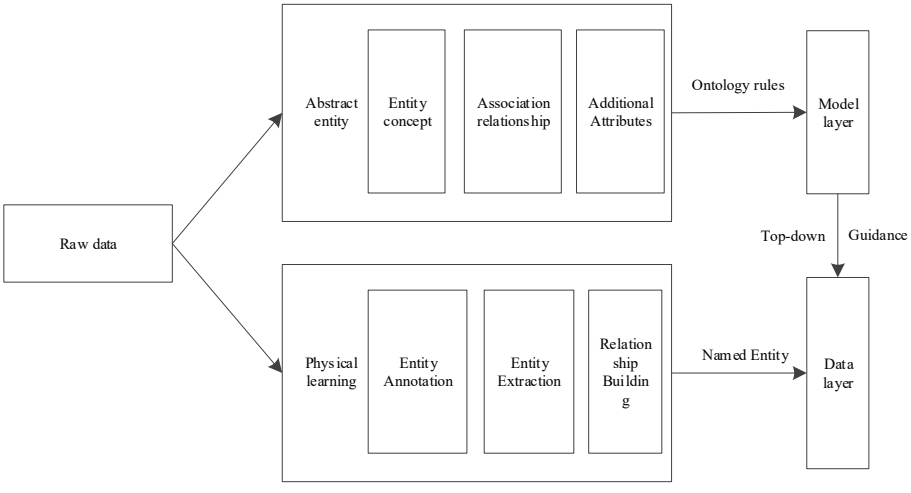
3 BiLSTM layer: pass the vector sequence T generated by the BERT layer into the BiLSTM layer to capture the contextual language features within the text, and finally output the fused context vector sequence $O = [o_1, o_2, \ldots, o_n]$.

4 Concatenate feature concatenation layer: firstly, the feature fusion process concatenates the vector sequences T and O generated by BERT and BiLSTM into a new output sequence $T_{out} = [e_1, e_2, \ldots, e_n]$. Subsequently, the sequence is fed into the fully connected layer to calculate the score of each character corresponding to various entity labels, and the final output score sequence is S.

5 CRF layer: finally, the score sequence S is fed into the CRF layer. This layer will automatically learn and update a state transition matrix with dimension $[n_i, n_j]$, calculate the loss function based on the score sequence and transition matrix, and finally output the optimal entity label sequence.

The construction process of a knowledge graph for distribution network fault handling is shown in Figure 2.

Figure 2 Construction process of knowledge graph for distribution network fault handling



Firstly, based on the structural characteristics of each type of data, design corresponding conceptual frameworks and association rules to construct the pattern layer; guided by the pattern layer, entities are extracted from unstructured text data and mapped and filled to form the data layer.

The knowledge graph update for automatic decision-making of distribution network fault handling is divided into two levels: the pattern layer and the data layer. The pattern layer takes distribution network equipment as the entity type, and its core content includes scheduling regulations and inter entity relationships. Its update triggering methods are divided into updating during regulation revision and manually adding settings (Xiong et al., 2026; Hou et al., 2025). The data layer includes specific data entities such as ledger data, fault handling records, and distribution network defect libraries, which are updated according to different cycles such as annual updates, semi annual updates, and quarterly updates. The knowledge extraction stage is responsible for

forming triplets based on the link relationship between the data and the pattern layer, and ultimately completing the continuous iterative update of the knowledge graph through the update process.

The knowledge map for distribution network fault handling is constructed from top to bottom, and its detailed structure is divided into two levels: mode layer and data layer. The mode layer takes distribution equipment as the core entity type, covering equipment entities such as transformers, feeders, switches, buses, distributed power sources and loads, and defines the operation rules, protection coordination relations, transfer path constraints and topological connection relations among the equipment in the dispatching regulations as the main relationship types, and sets the attribute fields such as equipment voltage level, operation state and rated capacity, forming the conceptual framework and association rule system of fault handling reasoning. The data layer is based on the constraints of the mode layer, mapping and filling the concrete entity instances and attribute values in unstructured and semi-structured ledger data, fault handling records, distribution network defect database and real-time operation data to form concrete knowledge instances stored in the form of triples, including historical fault modes of specific equipment, records of protection actions, success of reclosing and operation cases of load transfer, thus supporting subsequent reasoning tasks such as fault type discrimination, fault section location and recovery path search. The updating trigger mechanism of the knowledge map is set in the mode layer and the data layer respectively: the updating of the mode layer is triggered by the revision of dispatching regulations or the manual addition of new equipment types and new relationship definitions, and when the superior dispatching department issues new operation regulations or the distribution network structure undergoes major transformation, the updating process of the mode layer is started after being audited by professionals; the update of the data layer is automatically triggered according to a fixed period, in which the ledger data is updated annually, the fault handling records are updated semi-annually, and the distribution network defect database is updated quarterly. At the same time, when real-time fault events or protection action signals are detected, the system immediately triggers the incremental update of corresponding fault cases, and the newly occurred fault information is supplemented to the data layer in the form of triples after knowledge extraction. Compared with the traditional reinforcement learning method, this two-level updating mechanism has obvious advantages: traditional reinforcement learning relies on fixed state space and action space, and its Q-value function or strategy network cannot adapt to the topology change or regulation adjustment of distribution network after training. Once the grid structure or operation rules change, it is time-consuming and difficult to realise engineering. The updating mechanism of knowledge map can reflect the revision of regulations and the change of network structure to the mode layer in real time, and continuously enrich the new fault cases to the data layer, and strengthen the learning agent to obtain the latest topological relationship, equipment state and feasible operation set by searching knowledge map during decision-making, so as to dynamically adapt to the change of operation mode without retraining network parameters, effectively solving the problems of insufficient generalisation ability of traditional methods and the decline of decision accuracy in small sample scenarios, avoiding the decision delay caused by retraining, and improving the timeliness and adaptability of automatic decision-making.

The knowledge graph has completed the structured modelling of static rules, topological relationships, and historical cases involved in distribution network fault handling. However, the graph itself lacks autonomous decision-making capabilities, serving instead to provide prior knowledge constraints and state feature enhancements for decision-making algorithms. To achieve the automatic generation of fault handling action sequences from fault occurrence, it is necessary to introduce reinforcement learning algorithms capable of optimising decision strategies through interaction with historical experience in dynamic environments. Section 2.3 will elaborate on how to integrate the knowledge graph into the deep Q-network framework to achieve the deep fusion of knowledge and decision-making.

2.3 Automatic decision-making for distribution network fault handling based on reinforcement learning

2.3.1 Design a reinforcement learning decision framework that integrates knowledge graphs

Deep Q-network (DQN) is a representative reinforcement learning algorithm for optimising discrete sequence actions. This algorithm integrates Q-learning methods with deep neural networks and essentially belongs to the off policy class of algorithms. DQN contains two distinct strategies: the target policy and the behaviour policy. Among them, the target strategy π is responsible for learning and iteratively updating its parameters from historical experience, and tends to select actions with higher action value function values in action selection. This process can be expressed by the following relationship:

$$\pi(S_{i+1}) = \arg\max_{a_{i+1}} Q(S_{i+1}, a_{i+1}) \tag{7}$$

In the above formula, S_{i+1} represents the environmental state that the agent is in during the $i + 1$ iterations; a_{i+1} represents a specific action among all possible actions that the intelligent agent can take in state S_{i+1}; $Q(S_{i+1}, a_{i+1})$ represents the action value function.

The update method of the action value function of DQN algorithm can adopt the time series differential learning method with high update efficiency, and its calculation formula is as follows:

$$Q(s_t, a_t) \leftarrow Q(s_t, a_t) + \alpha\left[r_{t+1} + \gamma \max_{a_{t+1}} Q(s_{i+1}, a_{t+1}) - Q(s_t, a_t)\right] \tag{8}$$

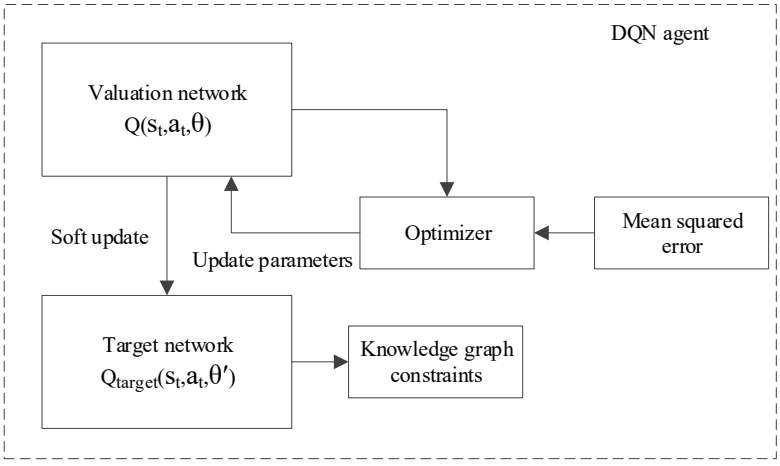
In the Q-learning algorithm, the action state value function is stored in a table format, which leads to an inherent flaw: when the problem scale expands and the number of states and actions increases, both the storage and reading of value data will consume significant storage resources and time costs, and may even cause dimension explosion problems. To overcome this deficiency, the DQN algorithm introduces a neural network to approximate the value of Q, and uses the network to fit the state action value function $S(s_t, a_t)$, that is

$$Q(s_t, a_t) = E_\pi\left[\sum_{k=1}^{+\infty} \gamma^k r_{k+1} \middle| s_t, a_t\right] \approx Q(s_t, a_t, \theta) \tag{9}$$

In the above formula, γ represents the discount factor, r_{k+t} represents the immediate reward value after the k^{th} action, and θ is the neural network parameter; $E_\pi(\cdot)$represents the expectation for strategy π.

In order to effectively converge the neural network, the DQN algorithm sets the target network with a parameter of θ. The reinforcement learning decision framework that integrates knowledge graphs is shown in Figure 3.

Figure 3 Reinforcement learning decision framework integrating knowledge graph



The motivation of adopting deep Q-network stems from the fact that the decision-making of distribution network fault handling is essentially a large-scale discrete sequence action optimisation problem. Its state space contains multi-dimensional continuous and discrete mixed variables such as voltage, current, power, switch state and topological connection relationship of each node, and the action space involves the combination of switching operation sequences of multiple switches. If the traditional Q-learning method is used to store the state-action value function in table form, it will face the problem of huge storage resource consumption or even dimension explosion, and the deep Q-network effectively overcomes this defect by fitting the value function with neural network. At the same time, the separation mechanism between the target network and the evaluation network and the experience playback technology introduced by it can break the data correlation, improve the training stability and convergence efficiency, making it especially suitable for sequential decision-making tasks such as distribution network fault handling with high real-time requirements and huge state space. The main innovations of this paper, which are different from the existing general research of deep Q network, are embodied in three aspects: first, the structural knowledge of knowledge map is integrated on the basis of traditional state input, and the topological relationship, historical fault mode and protection coordination rules of distribution network equipment are embedded in the map and spliced with real-time measurement data as the state input of agents, so that agents can understand the global topological semantics of distribution network instead of relying only on numerical characteristics, which is different from the practice of traditional DQN that only inputs the original sensing data directly into the network; secondly, a reward function with knowledge map is designed for the distribution network fault handling scenario, and the correctness of fault isolation, the recovery rate of non-fault sections, the punishment of operation steps and the constraint of equipment operation priority are simultaneously included in the instant reward calculation, so that the strategy learning process is highly consistent with the requirements of dispatching regulations, and the

limitation that the traditional reinforcement learning reward function is biased towards decision speed and ignores accuracy is overcome; thirdly, the target network parameters are copied from the evaluation network every fixed number of steps instead of updating every step, and the loss function is calculated and gradient descent is performed in combination with the time sequence difference error, so that the end-to-end dynamic generation of the disposal action sequence is realised in the specific task of distribution network fault disposal.

The deep Q-network used in this article is not a simple application of the standard DQN algorithm, but has been improved in three aspects: state space construction, reward function design, and network structure adaptation. At the level of state space construction, standard DQN usually directly takes sensor measurement data as state input and cannot perceive the topology connection relationship and logical constraints between devices in the distribution network. In this paper, based on the original measurement state, graph embedding representation of knowledge graph is introduced. By graph convolution encoding all relevant device entities and their topology adjacency relationships within the current fault impact range in the knowledge graph, the embedding vector is obtained, and the embedding is concatenated with real-time measurement data as the complete state input of the agent. This enables the agent to simultaneously perceive changes in measurement values and topological semantic constraints when making decisions, overcoming the lack of perception ability of standard DQN for network structure changes. At the level of reward function design, standard DQN, the reward for fault isolation is usually set based solely on the achievement or failure of the goal, resulting in low exploration efficiency in the early stages of training and a tendency for the strategy to lean towards decision speed while neglecting accuracy. This article decomposes the scheduling requirements for distribution network fault handling into multiple objectives and sets a dense reward function. The correctness of fault isolation, non-fault section recovery rate, operation step penalty, and equipment operation priority constraints are all included in the immediate reward calculation. The fault isolation correctness reward ensures that the intelligent agent accurately determines the fault location first, and the non-fault section recovery rate reward provides continuous value feedback according to the actual recovery load ratio. The operation step penalty encourages streamlined operations to reduce the number of switch actions. The equipment operation priority reward applies a large negative reward when the action sequence violates the protection cooperation relationship or the switching operation sequence, this dense reward mechanism enables the agent to obtain effective feedback signals at every step, significantly accelerating policy convergence and ensuring high consistency between decision results and scheduling regulations. At the level of network structure adaptation, the standard DQN adopts a uniform random sampling experience replay strategy. However, this paper designs a priority experience replay mechanism to address the uneven distribution characteristics of various types of faults in distribution network fault scenarios, where the frequency of occurrence varies greatly. Different sampling weights are assigned to the experience samples based on the absolute value of the temporal difference error, allowing the network to learn rarer but critical complex fault cases during the training process. At the same time, the parameters of the target network are not updated every step, but are synchronised from the estimation network using a phased replication strategy. This design reduces computational costs while avoiding training oscillations caused by frequent changes in target values, and improves the network's generalisation stability in diverse and non-fault scenarios. The above three

improvements together constitute a deep Q-network dedicated framework for distribution network fault handling tasks, which is fundamentally different from the general DQN algorithm and has stronger task specificity and engineering applicability based on the fusion of prior knowledge in the knowledge graph.

The above framework clarifies the combination of knowledge graph and deep Q-network, which uses graph embedding to combine topological semantics and measurement data as the state input of the agent, and designs a reward function that integrates regulatory constraints to guide strategy optimisation. However, this framework only specifies the value estimation method for single step decision-making, and further research is needed to provide a dynamic generation process for multi-step sequence decision-making, including specific update rules for the target network and valuation network, the construction method of the loss function, and the solution formula for the optimal action sequence. Section 2.3.2 will provide a complete derivation of the calculation process for sequence generation.

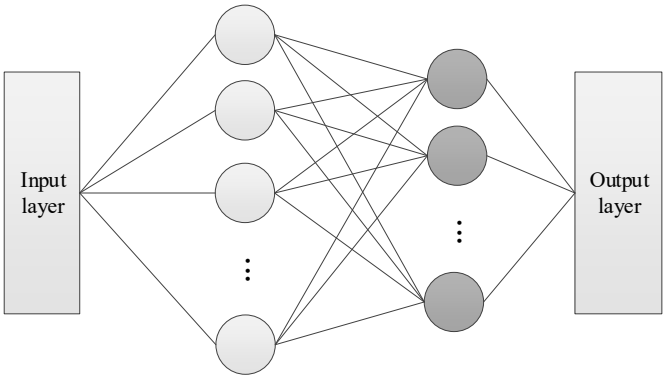
2.3.2 Dynamic generation of decision sequence for distribution network fault handling

Using the target network to calculate the state action function of the next state, indirectly obtaining the state action function value of the current state, and using it as the training target. In contrast, the parameters of the valuation network are updated in real-time, calculating the value function of the current state through different methods and minimising the difference between the target value, thereby optimising the neural network parameters of the valuation network. The calculation formula for Q values in the target network is as follows:

$$Q_{target}\left(s_t, a_t\right) = r_t + \gamma \max_{t+1} Q\left(s_{t+1}, a_{t+1}, \theta'\right) \tag{10}$$

To gradually fit the true state action value function of the neural network, the algorithm calculates the mean square error between the estimated network value $Q(s_{t+1}, a_{t+1}, \theta')$ and the target network output value $Q_{target}(s_t, a_t)$, and iteratively optimises the network parameter θ using this as the loss. The network adopts a fully connected architecture and consists of four layers: an input layer, two intermediate hidden layers, and an output layer. The schematic diagram of the neural network is shown in Figure 4.

Figure 4 Schematic diagram of neural network



The input layer of the intelligent agent receives $s_t = \{P_{LD}(t), P_{DG,fore}(t), G(t)\}$, and the output layer outputs the state action value, which estimates the network output of $Q(s_t, a_t, \theta)$. The target network obtains $\max_{a_{t+1}} Q(s_{t+1}, a_{t+1}, \theta')$, and then obtains $Q_{target}(s_i, a_t) = r_t + \gamma \max_{a_{t+1}} Q(s_{t+1}, a_{t+1}, \theta')$. The update strategy settings for the intelligent agent are as follows:

$$L(\theta) = E\left[\left(r_t + \gamma \max_{a_{t+1}} Q(s_{t+1}, a_{t+1}, \theta') - Q(s_t, a_t, \theta)\right)^2\right] \tag{11}$$

Then calculate the gradient of the parameters with respect to the loss function:

$$\frac{\partial L(\theta)}{\partial \theta} E\left[\left(r_t + \gamma \max_{a_{t+1}} Q(s_{t+1}, a_{t+1}, \theta') - Q(s_t, a_t, \theta)\right)\frac{\partial Q(s, a, \theta)}{\partial \theta}\right] \tag{12}$$

By continuously performing gradient descent to update parameter θ, the optimal state action value function can be obtained, and its calculation formula is as follows:

$$Q^*(s_{t+1}, a_{t+1}) = r(s_{t+1}, a_{t+1}) + \gamma \sum_{s' \in S} P(s' | s_{t+1}, a_{t+1}) \max_{a'} Q^*(s', a') \tag{13}$$

In the above formula, $r(s_{t+1}, a_{t+1})$ represents the immediate reward obtained by performing the action; $\sum_{s' \in S} P(s'|s_{t+1}, a_{t+1})$ represents the probability of state transition; $\max_{a'} Q^*(s', a')$ represents the maximum value of Q for the next state s'.

The calculation formula for dynamically generating the decision-making sequence for distribution network fault handling is as follows:

$$\tau^* = \arg\max_{\tau \in R} \sum_{t=0}^{T} \gamma Q^*(s_{t+1}, a_{t+1}) \tag{14}$$

At this point, this article has completed the complete method chain from data collection preprocessing, knowledge graph construction to deep Q-network sequence decision generation. The three links are progressive and mutually supportive: preprocessing provides high-quality data foundation for knowledge extraction, knowledge graph provides structured prior constraints for reinforcement learning, and reinforcement learning achieves automatic optimisation of fault handling action sequences under the joint drive of knowledge and data. To verify the effectiveness of the above method in actual distribution network fault scenarios, Section 3 will conduct experimental tests based on multi-source historical operating data and comprehensively compare it with two published methods.

3 Experimental design

3.1 Experimental protocol

In order to verify the practical application effect of the proposed method in automatic decision-making for distribution network fault handling, relevant tests were conducted. The specific experimental plan is as follows:

3.1.1 Experimental data

The collection of experimental data relies on multiple systems, including distribution automation system, data acquisition and monitoring control system, fault recording device, protection information management subsystem, and meteorological monitoring system. Among them, the data types that the distribution automation system can provide include: state variables of switches, action signals of protection devices, working status of fault indicators, online status of terminals, and telemetry data. The data acquisition and monitoring control system mainly outputs the following types of data: bus voltage value, line current value, active power, reactive power, and power factor; the data types provided by the fault recording device include voltage waveform at the fault moment, current waveform at the fault moment, harmonic components, and transient process records. The data types provided by the protection information management subsystem include protection settings, protection action events, fault distance measurement results, and reclosing action records. The data types provided by the meteorological monitoring system include wind speed, rainfall, lightning strike location, and temperature and humidity.

Preprocess the collected data: outlier detection and correction are based on the 3-fold standard deviation principle to identify values that deviate too much from the mean. Data with electrical quantity mutations exceeding 30% of the rated value are considered abnormal and replaced with sliding window averages. Time alignment resamples data with different sampling frequencies uniformly to the same time section, using nearest neighbour interpolation to maintain the original event order. Data filtering uses a second-order low-pass Butterworth filter to remove high-frequency noise interference, with a cutoff frequency set to ten times the power frequency. Data normalisation uses the maximum minimum value method to map electrical quantities to the range of [0, 1], while protection settings and meteorological data are normalised using standard scores. Data label annotation categorises samples based on the results of fault handling, including fault type, fault location, action correctness, and disposal timeliness evaluation.

The experimental data used in this paper are all from the historical operation data and fault cases recorded by actual operating systems such as distribution automation system, data acquisition and monitoring control system, fault recording device, protection information management subsystem and meteorological monitoring system, which belong to offline measurement data, that is, offline training and testing are carried out based on the records of real fault events that have occurred, so as to verify the historical retrospective decision-making effect of the method. In order to improve the network generalisation performance to cope with the diversity of distribution network operation modes and the randomness of fault scenarios, this paper introduces canopy clustering and UMAP manifold learning to preprocess high-dimensional operation data at the data level, preserves the global topological structure and reduces the redundancy of features, and at the same time uses the offline learning mechanism of random sampling in the experience playback pool to break the time series correlation of samples. In addition, the parameter separation structure of the target network and the estimation network itself also enhances the generalisation ability by reducing the variance of parameter update, so that the decision-making accuracy can still be maintained at a high level in the absence of fault scenarios.

3.1.2 Experimental indicators

Select the accuracy of automatic decision-making for distribution network fault handling, the time for automatic decision-making for distribution network fault handling, and the recovery rate of distribution network faults as experimental indicators to test the practical application effects of the method of Zhang et al. (2024) and method of Song et al. (2025).

The definition of the accuracy of automatic decision-making for distribution network fault handling refers to the percentage of times that the fault recovery power supply plan output by the automatic decision-making method after a fault occurs in the distribution network is consistent with the actual fault situation on site or the correct handling plan recognised by the dispatcher, which is a key criterion for measuring the core performance and reliability of the automatic decision-making method. It directly determines whether the method can replace or assist manual dispatchers in on-site operations. High accuracy automatic decision-making can significantly shorten the time of power outages caused by faults and reduce long-term power outages caused by human misjudgement or operation delays.

The automatic decision-making time for distribution network fault handling refers to the total time elapsed from the actual occurrence of faults in the distribution network to the completion of all disposal decisions by the automatic decision-making method and the output of the final operational instruction sequence that can be executed. This indicator directly determines the timeliness of fault handling and thus affects the reliability of power supply. Shortening the decision-making time can effectively reduce the duration of power outages for users and reduce the thermal shock damage of fault currents to equipment and lines. At the same time, rapid decision-making can avoid secondary accidents caused by the expansion of fault range due to delay. In addition, the stability of automatic decision-making time is also a key basis for judging whether the method meets real-time control requirements. A shorter decision-making time leaves sufficient time margin for subsequent remote control execution and switch actions, thereby improving the overall success rate of handling.

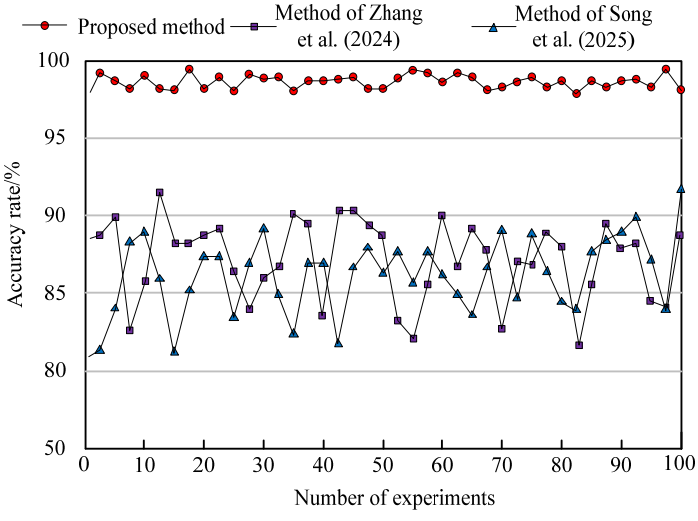
The recovery rate of distribution network faults refers to the percentage of non-fault power outage sections or loads successfully restored by the disposal plan generated by the automatic decision-making method after a fault occurs in the distribution network, compared to the total number of recoverable non-fault power outage sections or total loads. This indicator is a direct measure of the effectiveness of automatic decision-making methods in the non-fault section power supply recovery process, directly reflecting the self-healing ability and power supply reliability level of the distribution network. A high recovery rate means that the power outage range is minimised as much as possible during the fault handling process, and the power loss and social impact caused by the fault are minimised. At the same time, this indicator is also a key basis for evaluating the optimisation degree of the transfer strategy and the rationality of the network connection structure. Improving the recovery rate helps to reduce the average power outage time and frequency of users, thereby enhancing the continuous power supply guarantee ability of the distribution network for important loads.

3.2 Experimental results

3.2.1 Accuracy of automatic decision-making for distribution network fault handling

The comparison of the accuracy of automatic decision-making for distribution network fault handling using three methods is shown in Figure 5.

Figure 5 Comparison of automatic decision accuracy for distribution network fault handling (see online version for colours)



According to the analysis of Figure 5, the accuracy of the proposed method for automatic decision-making of distribution network fault handling is higher than the other two comparison methods in all experimental times, with a stable accuracy curve between 97.8% and 98.9%. The accuracy of the automatic decision-making of distribution network fault handling by method of Zhang et al. (2024) is between 81.8% and 91.7%, while the accuracy of the automatic decision-making of distribution network fault handling by method of Song et al. (2025) is in the range of 81.1% to 92.1%. These data indicate that the proposed automatic decision-making method is significantly superior to the two comparative methods in terms of fault location, isolation, and power restoration accuracy. It has better robustness and scene adaptability, and can maintain a high level of decision-making accuracy in large-scale repeated experiments or long-term operation. It has engineering practical value and promotion prospects.

3.2.2 Automatic decision-making time for distribution network fault handling

The comparison of automatic decision-making time for distribution network fault handling using three methods is shown in Table 2.

Table 2 Comparison of automatic decision time for distribution network fault handling

Number of experiments	*Proposed method (s)*	*Method of Zhang et al. (2024) (s)*	*Method of Song et al. (2025) (s)*
10	0.57	1.48	1.35
20	0.62	1.55	1.42
30	0.54	1.67	1.89
40	0.68	1.93	2.31
50	0.59	1.72	1.56
60	0.65	2.01	2.48
70	0.52	1.84	2.05
80	0.71	2.13	2.62
90	0.56	1.79	1.94
100	0.64	2.05	2.56

According to Table 2 analysis, the decision time of the proposed method is stable between 0.52 s and 0.71 s. The decision time range of method of Zhang et al. (2024) is 1.48 s to 2.13 s, while the decision time range of method of Song et al. (2025) is 1.35 s to 2.62 s. The time consumption of the proposed method is significantly lower than that of the two comparative methods. From the analysis of data volatility, the maximum and minimum values of the proposed method differ by 0.19 s, 0.65 s for method of Zhang et al. (2024), and 1.27 s for method of Song et al. (2025), indicating that the proposed method has better time stability under different experimental times. These data show that the proposed method has significant advantages in automatic decision-making speed for distribution network fault handling, mainly reflected in three aspects: firstly, it can complete decision calculations for fault location, isolation, and power restoration faster; secondly, it is insensitive to changes in experimental conditions or fault scenarios and has high robustness; the third is superior performance consistency, with no significant degradation or abnormal delay observed in up to 100 repeated experiments. In summary, the proposed methods effectively compress the calculation process by optimising the decision algorithm structure or data processing flow, achieving fast and stable automatic decision response capabilities, and are more suitable for application in distribution network fault handling scenarios with strict timeliness requirements.

3.2.3 Distribution network fault recovery rate

The comparison of the fault recovery rates of the three methods for distribution network is shown in Table 3.

Table 3 shows that the proposed method has a distribution network fault recovery rate between 96.78% and 99.12%, method of Zhang et al. (2024) between 87.56% and 92.13%, and method of Song et al. (2025) between 84.76% and 89.47%. The recovery rate of the proposed method was higher than the two comparison methods at all experimental times, with the lowest value of 96.78% still higher than the highest values of 92.13% in method of Zhang et al. (2024) and 89.47% in method of Song et al. (2025). From the fluctuation range, the range of the proposed method is 2.34%, the range of the method of Zhang et al. (2024) is 4.57%, and the range of the method of Song et al. (2025) is 4.71%. The stability of the proposed method is better. In 40 experiments, the proposed

method reached the highest value of 99.12%, with method of Zhang et al. (2024) at 92.13% and method of Song et al. (2025) at 89.47%, showing the most significant lead. These data indicate that the proposed method can restore power supply to more non-faulty power outage sections after a fault occurs. Its advantages are reflected in the high absolute recovery rate, significantly better performance lower limit than the upper limit of the comparative method, and more stable and reliable performance under different experimental times.

Table 3 Comparison of distribution network fault recovery rates

Number of experiments	*Proposed method (%)*	*Method of Zhang et al. (2024) (%)*	*Method of Song et al. (2025) (%)*
10	97.83	91.45	88.62
20	98.21	90.37	87.94
30	96.95	89.68	86.33
40	99.12	92.13	89.47
50	97.46	88.79	85.21
60	98.54	91.08	87.65
70	96.78	87.56	84.89
80	99.03	90.64	86.57
90	97.29	89.32	85.43
100	98.67	88.95	84.76

Through comprehensive comparison with the method of Zhang et al. (2024) and Song et al. (2025) in the above experiments and the method based on mixed integer programming mentioned in the introduction, the technical characteristics and core advantages of the proposed method can be summarised as follows: at the level of data preprocessing, in the proposed method, the strategy of combining canopy clustering with UMAP manifold learning is adopted. Canopy algorithm realises the rapid primary screening of data at the acquisition front end and has good tolerance for outliers. UMAP uses Riemannian geometry and algebraic topology to preserve the global topological structure in the process of dimensionality reduction of high-dimensional running data. This is different from the method of Zhang et al. (2024) which directly inputs the original features in the form of graphic data without targeted dimensionality reduction, and the BERT-BiGRU-CRF model in the method of Song et al. (2025) which only focuses on text knowledge extraction and ignores the preprocessing of numerical time series data. The dual preprocessing mechanism makes the proposed method more scalable and structural fidelity in the face of massive multi-source heterogeneous data. On the knowledge modelling level, the proposed method constructs a top-down two-level knowledge map, in which the mode layer encapsulates the scheduling rules and topological relation rules, and the data layer continuously enriches the account, defect and fault case entities, and sets up a separation mechanism that the regulation revision triggers the mode layer update and periodically triggers the data layer update. Compared with the method of Song et al. (2025), which uses secondary to realise the map visualisation but does not specify the two-level update strategy, this method can respond to the change of distribution network operation mode and regulation adjustment in time, and avoid the decision deviation caused by outdated knowledge. However, the method

based on mixed integer programming does not introduce the knowledge map mechanism at all, and can not effectively use unstructured fault text and scheduling regulation information. On the decision-making level, the proposed method designs a deep Q-network framework with knowledge map, embeds topological semantics into the state space and discipline constraints into the reward function, and automatically generates the optimal action sequence by using the offline learning strategy of separating the parameters of the target network and the estimated network, which is different from the method of Zhang et al. (2024) which only relies on the graph neural network to process the graph data input and the reward function is biased towards speed. The proposed method makes up for the deficiency of network generalisation in small sample scenarios through the global topological information and historical fault modes provided by the knowledge map. Compared with the problem of additional serial delay caused by knowledge map retrieval and reasoning in method of Song et al. (2025), the proposed method fuses knowledge map features and measurement data and inputs them into neural network for forward calculation at one time, thus avoiding the time overhead caused by multi-stage serial processing. Compared with the method based on mixed integer programming, the proposed method does not need to simplify the topology and does not adopt the conservative strategy of giving up the edge load, so it can maintain a high recovery rate and have stronger real-time response ability. The above technical characteristics jointly support the core advantages of the proposed method in the experiment: the decision-making accuracy is stable between 97.8% and 98.9%, which is significantly higher than that of the comparison method between 81.1% and 92.1%; the decision-making time is controlled between 0.52 s and 0.71 s, which is only about one-third of the time spent by the comparison method and the fluctuation range is the smallest; the failure recovery rate reaches 96.78% to 99.12%, and its lower limit is still higher than the upper limit of the comparison method, which comprehensively verifies the overall superiority of the proposed method in three dimensions: accuracy, timeliness and power supply recovery ability.

4 Conclusions

A new automatic decision-making method for distribution network fault handling based on knowledge graph and reinforcement learning is proposed to address the practical challenges of large information processing volume, strong decision-making timeliness, and variable operation modes in automatic decision-making for distribution network fault handling. Around the data collection and preprocessing stage of distribution network operation, canopy clustering algorithm is introduced to achieve rapid initial screening of data and UMAP manifold learning is used to preserve high-dimensional global topology structure. At the level of knowledge modelling, the BERT BiLSTM CRF model was used to extract entities and relationships from unstructured and semi-structured scheduling rules, fault handling records, and defect libraries. A top-down knowledge graph for distribution network fault handling was constructed, and a two-level update mechanism was designed for the pattern layer and data layer to ensure the timeliness and completeness of knowledge. At the level of decision generation, a deep Q-network framework that integrates knowledge graphs was designed. By separating the parameters of the target network and the valuation network and implementing a phased replication strategy, combined with temporal differential learning methods, the dynamic optimisation

generation of fault handling decision sequences was achieved. The experimental results confirm that the proposed method has significant advantages in decision accuracy with minimal fluctuation range, significantly faster decision time than the two comparison methods, and the lower limit of fault recovery rate is still higher than the upper limit of the comparison method, reflecting the comprehensive superiority of the proposed method in accuracy, timeliness, and power supply recovery capability. Subsequent research will focus on addressing multiple fault chain scenarios under extreme weather conditions and multi time scale collaborative optimisation decision-making problems, further enhancing the robustness and adaptability of the algorithm.

Acknowledgements

This work is supported by Guangzhou Power Supply Bureau Production Technology Renovation Project: Installation of Automatic Operation Function for Fault Handling in Edge Cluster Distribution Network Dispatching of the Control Center, grant no. 0301002025030301DD0013.

Declarations

All authors declare that they have no conflicts of interest.

References

Al-Jumaili, A.H.A., Muniyandi, R.C., Hasan, M.K., Singh, M.J., Paw, J.K.S. and Al-Jumaily, A. (2024) 'Parallel power load abnormalities detection using fast density peak clustering with a hybrid canopy-K-means algorithm', *Intelligent Data Analysis*, Vol. 28, No. 5, pp.1321–1346.

Bera, A. and Mande, S. (2025) 'Modeling and optimization of power distribution network using response surface methodology for power integrity analysis', *International Journal of Electronics and Communications*, Vol. 190, No. 1, pp.1–10.

He, S., Zhou, Y., Yang, Y., Li, X., Liu, Y. and Liu, T. (2025) 'Vulnerable sequence identification for sequential cascading failure analysis in power grid', *IEEE Transactions on Industrial Informatics*, Vol. 21, No. 6, pp.4830–4840.

Hou, B., Wang, Q. and Zhou, Z. (2025) 'Intelligent retrieval method for power grid dispatching information based on knowledge graph', *International Journal of Business Intelligence and Data Mining*, Vol. 26, Nos. 3–4, pp.431–447.

Li, A. and Chauhan, A. (2025) 'Analysis of human focus patterns using uniform manifold approximation and projection', *International Journal of High School Research*, Vol. 7, No. 9, pp.12–19.

Li, B., Zheng, Y. and Ran, R. (2025a) '2DUMAP: two-dimensional uniform manifold approximation and projection for fault diagnosis', *IEEE Access*, Vol. 13, No. 1, pp.12819–12831.

Li, T., Wu, Z., Liu, Y. and Lin, Q. (2025b) 'Research on emergency repair of distribution network considering cooperative scheduling of power supply vehicles', *IEEJ Transactions on Electrical and Electronic Engineering*, Vol. 20, No. 3, pp.349–357.

Liu, M., Du, M. and Mu, W. (2025) 'Power grid engineering knowledge graph and physical elements alignment algorithm designs', *Annals of Data Science*, Vol. 12, No. 6, pp.2009–2039.

Song, D., Yuan, L., Xiao, X.T., Du, X.X., Shang, X.X. and Yang, Z.H. (2025) 'Research and application of knowledge graph for fault assisted decision-making in distribution network based on BERT-BiGRU-CRF', *China Electrical Equipment Industry*, Vol. 12, No. 4, pp.23–28.

Wang, Y. and Sun, N. (2025) 'Intelligent early warning method of power grid failure events based on temporal convolutional neural networks and rule learning', *Journal of Physics: Conference Series*, Vol. 2936, No. 1, pp.012037–012048.

Xia, T., Shen, X.W. and Shang, Y.W. (2024) 'Power distribution system multi-objective maintenance scheduling based on mixed-integer programming and its sensitivity analysis', *Power System Technology*, Vol. 48, No. 11, pp.4680–4689.

Xiong, J., Yang, P., Chen, B. and Chen, Z. (2026) 'Defect identification method of power grid secondary equipment based on coordination of knowledge graph and Bayesian network fusion', *Energy Engineering*, Vol. 123, No. 1, pp.296–313.

Zhang, P., Chen, Y.X., Wang, G.H. and Li, X.Y. (2024) 'Fault recovery decision of distribution network based on graph reinforcement learning', *Automation of Electric Power Systems*, Vol. 48, No. 2, pp.151–158.

Zhang, R. (2024) 'Improve threshold range of canopy clustering using optimization algorithms', *Asian Journal of Research in Computer Science*, Vol. 17, No. 12, pp.148–164.

Zhou, T., Li, X. and Lu, H. (2026) 'A dual power grid cascading failure model for the vulnerability analysis', *IEEE Transactions on Smart Grid*, Vol. 17, No. 1, pp.590–602.