

¹Zhongqiang Zhou¹Xiaofang Liu¹Yingjie Li¹Junlin Li²Xuanlin Long²Jupeng Zeng

An Automation Terminal Optimal Configuration Method Based on Knowledge Graph for New Distribution Networks



Abstract: - To ensure the safety and dependability of the new distribution network, the need for distribution automation terminals develops quickly, which contributes to the high cost of protection and building of distribution network. In this paper, a new method for optimizing the automation terminal configuration based on knowledge graph is proposed. In this method, redundant lines are calculated by establishing knowledge graph of distribution network, and the automation terminal of distribution network is rationally allocated. The input of the automation terminal is lowered, and the cost is reduced on the premise of assuring the reliability and sensitivity of distribution network protection. The approach first creates the knowledge graph of distribution network, which can give accurate data and clear topology diagram of power network and puts forward the corresponding judgment foundation. On this basis, the complicated distribution network is layered, and the constraint requirements are established in layers. The redundant lines are obtained using improved simulated annealing particle swarm optimization. Finally, the simulation and laboratory tests show that the proposed method can effectively avoid the impact of new energy access on the optimal configuration of distribution network automation terminals. The optimal configuration can ensure the reliability of fault response and the accuracy of fault location, and effectively reduce the cost of protection configuration in the new distribution network with distributed generation.

Keywords: New distribution network, automation terminal, optimal configuration, knowledge graph.

I. INTRODUCTION

Distribution network is intimately associated with users, is the final link of power transmission, impacting all areas of everyday production and living, thus the power supply dependability of distribution network is of considerable relevance [1]. With the acceleration of the building of new power systems, various distributed generations (DG) such as solar energy, wind energy and electric cars are connected to the grid. Due to its limited capacity and low grid-connected voltage level, DG is more ideal for connecting to the distribution network [2]. However, when the distributed generation is linked, the power flow direction of the traditional distribution network will change, which will present issues to the protection identification, operation, fault detection and other elements of the distribution network [3].

With the progress of high-precision clock synchronization and power system communication technology, the traveling wave placement technique based on the full power grid has been created and has become a research hotspot [4-6]. These strategies tackle the previous difficulties of single source and low dependability of traveling wave location data and present a new notion for traveling wave location. However, due to the huge number of automation terminals, high investment cost and long construction time, it becomes a challenging challenge for the continued development and deployment of traveling wave location in power grid. To overcome this difficulty, academics have investigated the ideal configuration approach of automation terminals. Reference [7] established the goal function of the optimization model by incorporating the involvement of the full fault monitoring unit and the uniform setup of the monitoring equipment. Combined with the economic limitations, the ideal configuration model of fault finding device for low voltage new distribution network is created. Reference [8] expressed the topology of distribution network based on graph theory and established the optimal configuration model of distributed intelligent terminals with the goal of optimizing the comprehensive calculation load of distributed intelligent terminals in the distribution network area. The model solving approach based on Lagrange multiplier method is proposed. Reference [9] employed binary particle swarm optimization method to tackle the problem of distribution network failure indication optimization and compared the outcome with immune algorithm. Reference [10] presented the optimum allocation model of FTU with the purpose of maximizing the observable rate of distribution

¹ Power Dispatching Control Center, Guizhou Power Grid Co.,Ltd, Guiyang, 550002, China

² State Key Laboratory of Disaster Prevention and Reduction for Power Grid, Changsha University of Science and Technology, Changsha, 410114, China

* Corresponding Author: Jupeng Zeng. Email: zengjupeng1224@stu.csust.edu.cn

network and bound by the reliability and economy of distribution network power supply. Reference [11] provided an optimal configuration model of power acquisition terminal of charging pile network based on integer linear programming. However, there are several unclear elements in microgrid system, such as parameter disturbance and input disturbance, which need additional investigation.

As an essential field of artificial intelligence technology, the successful application of knowledge graph in big data environment offers humans with a more effective trans-media big data organization, management, and cognitive capacity [12]. The knowledge graph also demonstrates remarkable strength in intelligent search, intelligent question and answer, intelligent recommendation, and intelligent decision-making applications paired with large data [13]. The domain-specific knowledge graph (DKG) is a knowledge representation that translates the data information in a given domain into a knowledge representation that is closer to the human cognitive environment [14]. In distribution protection system, knowledge graph technology can clearly examine the link between things. It may complement the real information, determine the validity of the existing information, and dig out the hidden internal connections in the system through the mining and reasoning of the inter-entity relationship [15-17]. The qualities of knowledge graph fully fulfill the requirements of topology identification and decrease the dependency of topology identification on data. According to the data characteristics and development trend of electric power Internet of Things of our nation, Reference [18] put up a design of application framework for knowledge graph of the electric power domain based on NoDKG(not just domain-specific knowledge graph). The application direction and future problems of domain-specific knowledge graph are outlined and prospected. Reference [19] integrated knowledge graph technology into low-voltage distribution network and applied it to the topology identification of low-voltage distribution network for the first time. Reference [20] suggested an auxiliary decision approach for distribution network fault scheduling based on knowledge graph. This method considers the actual scenario of the control operation task, constructs the knowledge graph of distribution network fault scheduling, and studies the application idea and implementation method of interactive reasoning decision making based on business requirements. Reference [21] proposed a combination of top-down and bottom-up knowledge graph construction method and proposed a set of knowledge extraction method based on deep learning in the power field according to the characteristics of the plan text. The preceding literature has effectively used knowledge graph technology to development planning, operation and maintenance, power grid dispatching, equipment consultation and quality control in the power industry. But nearly no academics have utilized knowledge graph technology to the optimization of distribution network automation terminal setup.

Aiming at the radiant topology of the new distribution network, based on the top-down knowledge graph construction method and the double-ended traveling wave location principle, this paper completes the automation terminal optimal configuration of the new distribution network. The knowledge graph firstly constructs a schema layer, which takes the minimum number of automated terminals as the objective function, and the simulated annealing particle swarm optimization algorithm is used to solve the problem under the constraint condition that the whole network faults can be observed. Based on the analysis of distinct fault positions and types, the fundamental conditions of terminal setup are summarized, and the complex modeling of observable constraint conditions for the full network is avoided. The weight coefficient is included to quantify the real demand of each node on the automation terminals so that the design scheme is more practical. Both simulation and laboratory tests indicate that the proposed knowledge graph layout optimization technique can install the least number of automation terminals, and assure that any line defect can be seen in the full network. The proposed method can effectively avoid the impact of new energy access on the optimal configuration of distribution network automation terminals. Moreover, this result can also be utilized for fault location of indirectly observable lines, and the fault location error is smaller than 100m.

Wherever Times is specified, Times Roman or Times New Roman may be used. If neither is available on your word processor, please use the font closest in appearance to Times. Avoid using bit-mapped fonts if possible. True-Type 1 or Open Type fonts are preferred. Please embed symbol fonts, as well, for math, etc.

II. FAULT TRAVELING WAVE TESTABILITY STUDY OF DG ACCESS POINT

There are several ways for distributed generation to connect to the grid, mostly governed by the capacity of the distributed generation and the access voltage level of the distributed generation. Since the capacity of most distributed generation is not too great, the voltage level of access to the distribution network is low, and the position of access points gives the potential for the installation of automated terminals, thus it is important to examine the access mode. According to the varying grid-connected voltage and capacity, distributed generation may be

separated into small, medium, and big grid-connected power station. The voltage level of the tiny grid-connected power station is 0.4kV low-voltage power station. The entire installed capacity is not less than 50kW, often not more than 200kW. Medium grid-connected power stations refer to new energy power stations having access voltage ranges from 10 to 35kV. The total installed capacity is commonly classified into three categories: A, B and C, as shown in Table 1. Large grid-connected power stations refer to power stations having access voltage levels of 110kV and higher, and the total capacity is often larger than 20000kW.

Table 1 Classification of DG

Classification	Capacity/kW	Grid-connected voltage level
Medium station (A)	$6000 \leq P \leq 20000$	35kV
Medium station (B)	$1000 \leq P \leq 6000$	10kV
Medium station (C)	$200 \leq P \leq 1000$	10kV
Small station	$50 \leq P \leq 200$	400v

According to the different grid-connection techniques, it can be largely split into substations linked to the distribution network through special lines, T connected to the distribution network lines, and access to the user's own power supply system, as shown in Fig. 1. It is mainly divided into access to 380/220V low-voltage lines of the distribution network, access to 380/220V user systems, and the rest can be accessed online; It is connected to 10kV line or substation by cable line, connected to 10kV line of distribution network and connected to 10kV user system, and the remaining part is connected to the Internet.

According to the transmission characteristics of the distribution transformer to the traveling wave and the propagation characteristics of the traveling wave signal in the distribution line, it can be seen that the traveling wave signal may be difficult to detect if the access point of the distributed generation accessing the 380/220V voltage class is located on the low voltage sideline of the distribution transformer. Therefore, this article mainly concerns access points with access voltage class of 10kV.

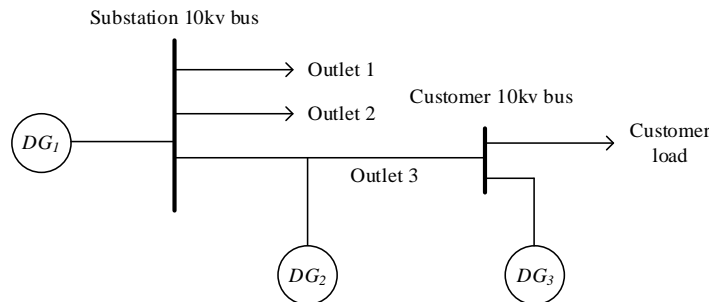


Figure 1 DG access mode

Based on the examination of the access locations of the above three access modes, the situation of DG1 accessing the bus bar of the substation is not mentioned since the automation terminals may be put in the substation. For the situation where T of DG2 is linked to the distribution line, it can be similar to the case of a simple branch on the distribution line, and traveling wave signals can be recognized at the junction point. For DG3, it relies on the number of outgoing lines on the user's 10kV bus bar. Generally, if the number of users is not big, the voltage traveling wave signal may also be recognized at the parallel point. However, if there are several lines, it is suggested not to install a automation terminals here.

III. CONSTRUCTION OF KNOWLEDGE GRAPH SCHEMA LAYER

In the application of knowledge graph, the knowledge graph has nodes, edges and attributes and leverages the structure of distribution graph to conduct semantic query of distribution network database. In the knowledge graph, each node or edge represents an entity, and the relationships between these items are represented by the graph structure. Graph database has become a strong tool for expressing entities because it can flexibly hold complicated relationships between items. In modeling, "edge" is used to explicitly describe the link between entity records. When querying, entity records may be accessed directly by leveraging the relationship defined by "edge", which reduces the calculation cost of standard relational database processing "table association". In this research, a top-down strategy is utilized to develop knowledge graph. Firstly, according to the structural features of different forms of data, the matching conceptual model and rule connection are developed, and the schema layer is produced. Then,

depending on the schema layer, entities are retrieved from the text data to generate the associated data layer. The specific building procedure is presented in Fig. 2.

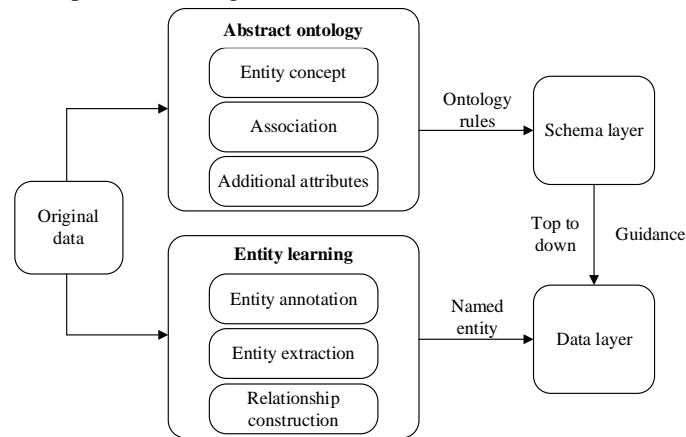


Figure 2: Construction process of knowledge graph

The objective of abstract ontology is to find the set of tags used to represent concepts, entities or attributes, and extract entity concepts, association relations and additional attributes through various types of data to form corresponding ontology rules to build the schema layer of knowledge graph. The heart of knowledge graph of optimal configuration is optimization rules. The primary goal of things such as nodes and lines are to concretely represent new distribution network elements commonly referred to in optimal configuration. Therefore, the schema layer design of knowledge graph of optimal configuration should focus on optimization rules.

3.1 Automation terminal layout optimization rules

The theoretical study and implementation of traveling wave fault location method is becoming more and more developed. Because the new distribution network layout contains numerous branches, the multi-terminal fault location approach employing multi-point traveling wave head information is better suited for the new distribution network. If any line in the network is defective, the fault location or fault location may be established using the two-terminal location theory, and the fault line in the whole distribution network can be substantial. Therefore, the optimization criteria in this research are based on the traveling wave double-terminal ranging approach, which may split the new distribution network lines into two categories: directly observable lines and indirectly observable lines. directly observable lines: When a fault occurs on a line in the distribution network, if the automation terminals is installed at the end of the line or other locations, and it can ensure that the fault point is at least on the shortest path composed of two measurement points, the line can be said to be directly observable.

indirectly observable lines: If a line in the distribution network breaks, the signal created by the fault cannot be directly observed, that is, the fault location is not situated on the shortest path between the two sets of measurement stations. However, at least the number of nodes between a set of shortest paths in the network is 1, And the traveling wave signal created by the fault point can be detected by the measuring points at both ends of the shortest path, or the fault branch can be discovered by the combination of the data of each measuring point, then we name the line indirect observable.

Fig. 3 is the topology diagram of the complicated distribution network. In this study, each node is treated as "Vertex", and each portion of the line is numbered as "Edge". Using node A as the reference node, data in Table 2 can be obtained.

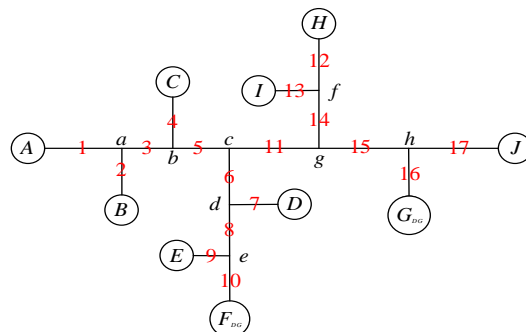


Figure 3: Active distribution network

Table 2: Relationship between Vertex and Edge

Vertex	Edge	Count
A,B	1,2	2
A,C	13,4	3
A,D	1,3,5,6,7	5
A,E	1,3,5,6,8,9	6
A,F	1,3,5,6,8,10	6
A,G	1,3,5,11,15,16	6
A,H	1,3,5,11,14,12	6
A,I	1,3,5,11,14,13	6
A,J	1,3,5,11,15,17	6

The distribution network is stratified, and based on the stratification results, the set of lines containing the most elements is defined as the first layer line of the distribution network and the main line of the distribution network. This work adopts A as the reference node, picks the path A-a-b-c-g-h-J in Fig. 3 as the first-layer line, A-J is defined as the first layer, and the set of lines included within is {1,3,5,11,15,17}. The intersection point (abcgh) passed by the first layer is defined as the starting point of the second layer, and the node set with the same trunk line elements in nodes except A and J in Table 2 is considered to be the same branch, and all the lower line node sets [B], [C], [D,E,F], [H,J] and [G] are obtained. As with the first layer line selection rule, the node that contains the most line elements in each collection is selected and defined as the end point of the second layer. And so on until the distribution lines are layered. As shown in Fig. 4, the power distribution circuit is divided into three layers according to the above principles. The first layer is selected as A-J, and the corresponding path is A-a-b-c-g-h-J. The second layer is: a-B, b-C, c-F, g-H, and the corresponding paths are a-B, b-C, c-d-e-F, g-f-H, h-G. The third layer is d-D, e-E, and f-I. The hierarchical structure of the distribution line section is shown in Fig. 4.

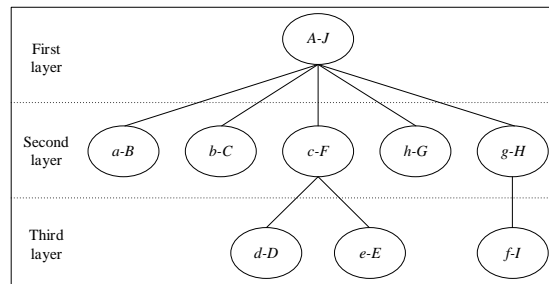


Figure 4: Hierarchical diagram of new distribution network structure

Consider the hierarchical structure diagram in Fig.4 and the constraints of directly and indirectly observable lines of the distribution network together. If the faults can be observed when faults occur on all layers of lines, then the entire network is measurable. The first layer line is defined as the trunk line. The line with no layers under a layer line is a single branch line, while the line with other layers is a multi-branch line. According to the concept of direct and indirect line fault observation, the automation terminal configuration of trunk line, single branch line and multi-branch line fault observation is evaluated separately.

3.1.1 Trunk line fault

It is characterized by the direct observation of the fault, and there are locating equipment at both ends of the fault site. Each segment of the trunk line is established as the fault line, and the installable nodes at both ends are separated into two arrays, as shown in Table 3. By taking the minimal common set of the array, the terminal configuration conditions for the direct observation of trunk line faults may be achieved.

Table 3: Node array on both sides of each section of trunk line

Faulty line	Left node array	Right node array
1	A	B,C,D,E,F,G,H,I,J
3	A,B	C,D,E,F,G,H,I,J
5	A,B,C,	D,E,F,G,H,I,J
11	A,B,C,D,E,F,	G,H,I,J

15	A,B,C,D,E,F,H,I	G,J
17	A,B,C,D,E,F,G,H,I	J

Table 3 reveals that the minimal common sets of the left and right node arrays are [A] and [J] correspondingly. Therefore, the automation terminal put at two nodes in the set [A,J] can achieve the direct observation of trunk line faults, and the number of terminals is minimal.

On the other hand, according to the concept of indirectly observable lines, when an automated terminal is built on a line that is one connection point distant from the two nodes of the trunk line, the accompanying end line is set as indirectly observable. For example, on the A-J trunk line in Fig.3 the line 1 and line 17 segments can be changed to indirect visibility by placing automation terminals at nodes B and G. At this moment, the fault may still be visible along the complete portion of trunk line A-J.

To sum up, if the automation terminal is installed at the first and last node in the first layer, or the point with the number of nodes between the first and last node is 1, all the fault points on the main line can be observable. The constraint relationship can be stated as:

$$f_1 + f_{1,1\pm 1} \geq 1; f_n + f_{n,n\pm 1} \geq 1 \tag{1}$$

In Eq. (1): f_1 is the top end of the main stem; $f_{1,1\pm 1}$ is the connection point with the number of nodes of 1 between the first endpoint; f_n main stem upper terminal point; $f_{n,n\pm 1}$ is the connection point with 1 node between the end point.

3.1.2 Single branch line fault

When a fault occurs on a simple branch, such as the line 2, or the line 4, the fault can be observed indirectly. Based on the two-ended traveling wave fault location principle, the measurement points installed on both sides of node b can locate the fault location on node b of the trunk line. In addition, the simple branch does not contain other branches, and the fault branch can be accurately located, meeting the indirect observable definition of the fault. Therefore, there is no need to deploy automation terminals for simple branches.

3.1.3 Single branch line fault

When a fault occurs on a multi-branch line, such as the c-F branch or other small branches of a large c-F branch, the fault traveling wave propagation will be more complicated and require specific analysis, as shown in Fig. 5:

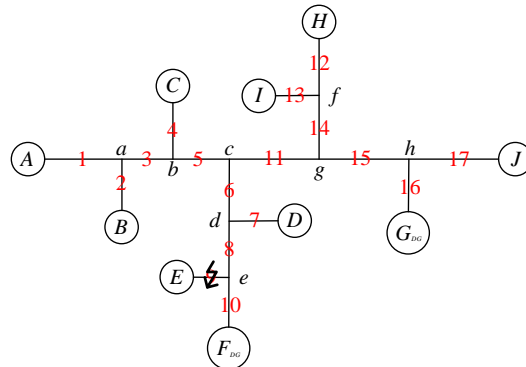


Figure 5: Multi-branch line fault

Assume that automated terminals are installed at A and J on the main line. If the line 9 fails, locate the fault on node c. As the cF branch is a multi-branch, the specific location of the fault point cannot be accurately located at this time, and specific analysis is required. Therefore, the corresponding point should be selected among the three points D, E and F to install the terminal. If the terminal is installed at point D, the line 7 fault is visible, but the line 8,9,10 fault is invisible; When the terminal is installed at point E, the failure at any point on the branch is visible; If the failure occurs on the cF trunk branch, the installation of automation terminal at point E can also ensure that the failure at each point is visible. The analysis shows that when the terminal is installed at point F, the installation effect is the same as that at point E, and the failure in any section on the branch is considerable. So, you can choose to install the terminal at any point in point E or F. However, the branch node between E and F is 1, so choose a location to install the automation terminal according to the weight coefficient. The same results can be obtained by analyzing other complex branches in the same way. Based on the above analysis, constraints are obtained:

$$\begin{cases} f_{n_1} + f_{n_1, n_1 \pm 1} \geq 1 \\ f_{n_2} + f_{n_2, n_2 \pm 1} \geq 1 \\ \dots\dots\dots \\ f_{n_i} + f_{n_i, n_i \pm 1} \geq 1 \end{cases} \tag{2}$$

In Eq. (2), f_{n_i} is the end node of multi-branch i of the distribution line; $f_{n_i, n_i \pm 1}$ is the connection point with the number of end nodes of multi-branch i of the distribution line being 1.

3.2 Set weight coefficient

As mentioned above, there are various schemes for the layout of automation terminals on trunk lines and multi-branch line. And these solutions are the same in terms of the number of terminals and the cost. To meet the objective function more effectively and improve the reliability of the corresponding nodes, it is also necessary to introduce the weight coefficient to select the optimal scheme.

3.2.1 Constraints on configured nodes

The installed automation terminal will not be removed in the actual project. If the automation terminal is installed on an existing node before the optimization of the network-wide automation terminal configuration, the optimization result must keep the installation status of the existing nodes. The following equality constraints are therefore attached.

$$f_p = 1, f_p \in s_{placed} \tag{3}$$

In Eq. (3), s_{placed} is the set of nodes on which an automation terminal has been installed before the optimal configuration.

3.2.2 Line length weight

Considering the different length of the line connected to the end nodes of each line, the attenuation degree of the fault traveling wave is different. In this paper, the weight of network line length is calculated to quantify the influence of line length on the configuration requirements of each node terminal.

$$L_i = \frac{l_i^2}{l'^2} \tag{4}$$

In Eq.(4), L_i is the line length weight; l_i is the shortest path length from node i to the middle point of the trunk line; l' is the average distance between the end node of each line and the midpoint of the trunk line.

3.2.3 Historical fault frequency weight of the line

To ensure the reliability of the power supply, the fault location should be determined in time for the line with a high historical failure rate, and the fault should be eliminated in the shortest time to restore the power supply. Therefore, this kind of line has a higher demand for the installation of automation terminals. In the actual optimization configuration, the historical fault frequency weight f_i of each line in the target distribution network can be obtained by sorting out the historical fault rate of each line in the network. In this paper, Matlab is used to generate the historical fault frequency array of line simulation, and the historical fault rate of each line terminal node is calculated by using the array.

3.2.4 Objective function weight

The different historical fault frequency and length of the corresponding line of each node result in the different requirements for the actual installation of the automation terminals. The difference can be reflected by introducing the weight coefficient into the objective function. The node weight coefficient w_i is obtained by combining the weight of line length and the weight of historical fault frequency.

$$w_i = C\left(\frac{1}{f_i} + L_i\right) \tag{5}$$

In Eq. (5), C is the weight coefficient adjustment factor, so that w_i is within $\{0,1\}$.

$$f(x) = \min \sum_{i=1}^N w_i x_i, i = 1, 2, \dots, N \quad (6)$$

In Eq. (6), $f(x)$ is the objective function after the introduction of the weight coefficient, and w_i represents the degree of automation terminal installation demand on each node. The smaller the value, the higher the degree of demand, so in the process of model solving, the automation terminal is more inclined to be configured on this node.

IV. CONSTRUCTION OF KNOWLEDGE GRAPH DATA LAYER

The construction of data layer is mainly divided into three steps: knowledge extraction, knowledge integration and knowledge updating. In this paper, knowledge extraction mainly refers to the entity recognition model under the guidance of schema layer to extract named entities from unstructured data and organize them into structured knowledge. knowledge integration refers to the integration of fragmented knowledge into usable knowledge. knowledge updating refers to the corresponding modification of knowledge graph according to the real-time changes of knowledge and new information following the building of knowledge graph.

4.1 Set weight coefficient

In this paper, the knowledge extraction is carried out based on the guidance of schema layer. Information is extracted, screened, and summarized from the original data, and ontology-based knowledge representation is formed on this basis. The distribution line and node data studied in this paper have highly structured characteristics. The corresponding link rules are designed under the guidance of schema layer, which can be used in the construction of knowledge graph. Some concrete examples are shown in Table 4.

Table 4: Sample of knowledge extraction

Entity category	Entity example
Line number	1,6,13
Line length	0.8km,1.8km,2.5km
Line wave impedance	300Ω,400Ω,500Ω
Automation terminal installation	Installed, not installed
Historical failure rate	0.04,0.18,0.24,0.61
Type of node	Customer load, wind power, thermal power
Capacity	0.8MW,2.4MW

4.2 Knowledge integration

Through knowledge extraction, we get a series of flat information which lacks the sense of hierarchy and logic. And knowledge integration is to sort out and merge, eliminate contradictions and ambiguities, and integrate knowledge fragments into usable knowledge. knowledge integration includes entity linking and knowledge merging. The entity link mainly refers to the entity extracted from the text data and the operation of linking it to the corresponding correct entity in the knowledge graph. Because of the highly structured characteristics of the extracted knowledge, there is no problem of entity disambiguation and co-reference resolution. Knowledge merging refers to integrating the existing relational databases of power system into the newly constructed knowledge graph through RDB2RDF (relational database to RDF) technology to further improve the total knowledge of knowledge graph when constructing knowledge graph.

4.3 Knowledge updating

With the iterative updating of time and data, new knowledge is increasing day by day, and the integrity of knowledge graph needs to be guaranteed by knowledge updating. According to logical architecture, the update of knowledge graph can be divided into schema layer update and data layer update. Schema layer update indicates that a new entity type appears in the newly added data and needs to manually update the new entity type to the schema layer. Data layer update refers to updating new data into the knowledge graph cell according to the previously established knowledge extraction rule under the guidance of the schema layer.

V. OPTIMAL CONFIGURATION BASED ON SIMULATED ANNEALING PARTICLE SWARM OPTIMIZATION

5.1 Simulated annealing particle swarm optimization

The idea of simulated annealing algorithm (SA) was first proposed by N.Metropolis et al. The starting point of its solution is to simulate the cooling process of solid temperature reduction and find out the similarity between the optimization problem and the cooling process to solve the optimization problem. SA has now developed into a more general approach to optimization problems, especially in engineering and management science. The simulated annealing algorithm simulates the cooling process of the classical particle system in thermodynamics to solve the optimal solution of the planning problem. When the solid temperature drops gradually, if the process is very slow, it can be approximated to that the object is in thermodynamic equilibrium. As the solid temperature progressively declines, the system temperature will also gradually fall and ultimately reach the lowest point, which corresponds to the optimization issue. PSO algorithm is an intelligent algorithm based on the activities of biological groups. PSO algorithm is simple, easy to calculate and implement, and has few parameters, fast convergence speed and easy parameter setting. Moreover, the algorithm converges faster in the early stage, but it will produce random oscillations in the later stage. If the search speed is slow near the optimal value, it may fall into the local optimal, the accuracy is reduced, and it is easy to scatter.

The approach based on the combination of simulated annealing algorithm and particle swarm optimization algorithm can solve the weaknesses of particle swarm optimization algorithm. It can greatly increase the operation speed, improve the related performance of the operation, and avoid falling into local optimization. Based on this idea, a particle swarm optimization algorithm based on simulated annealing process is established. The idea of simulated annealing is introduced into the particle swarm optimization algorithm, and the simulated annealing mechanism is introduced in the process of particle updating position and velocity. For the value following particle swarm development, the optimum solution is accepted according to a given probability, and the degraded solution is accepted to a certain amount. The annealing temperature is automatically adjusted, and finally the global optimal solution is found due to the decrease of particle energy. The optimization steps are as follows:

- (1) Initialize parameters: inertia weight; Acceleration constants $C1$ and $C2$; Annealing start and end temperature T and $T0$; Annealing rate K ;
- (2) The initial solution is given according to the actual distribution network.
- (3) Calculate the fitness value of each particle $f(Xi(t))$, ($i = 1, 2, 3, \dots, n$).
- (4) After calculating $f(Xi(t))$ for each particle through the fitness value function, they are compared with the individual extreme value, and the individual extreme value is updated after the optimization.
- (5) After calculating $f(Xi(t))$ for each particle through the fitness value function, they are compared with the global extreme value, and the global extreme value is updated after the optimization.
- (6) Update the position $Xi(t+1)$ and flight speed $Vi(t+1)$ of each particle according to the formula and limit the speed to the maximum speed V_{max} .
- (7) Calculate the updated adaptation value $f(Xi(t+1))$ of each particle.
- (8) Calculate the amount ΔE of change in the adaptation value caused by the two positions, if $\Delta E < 0$, accept the new position. If $EXP(-\Delta E / T) > \varepsilon$, ε is a random number between $\{0,1\}$, the new position is also accepted, otherwise rejected. If the new value is accepted, cool T , otherwise do not cool, return to step (3).

5.2 Optimization configuration scheme of automation terminal

In this paper, the constraints in the mathematical model of optimal configuration are dealt with by the non-fixed multi-segment mapping penalty function. The constrained optimization problem in the model is transformed into an unconstrained generalized objective function, whose form is shown below.

$$F(x) = f(x) + \delta(t)H(x) \quad (7)$$

In Eq.(7), $F(x)$ is the objective function in the model; $\delta(t)H(x)$ is the penalty term; $\delta(t)$ changes with the number of iterations. The penalty function can adaptively select different punishment intensity according to the degree of constraint violation, to avoid the difficult problem of selecting penalty factor when using precise penalty function. Then, the improved gray Wolf algorithm is used to intelligently optimize the generalized objective function, and the optimal solution of the problem is obtained.

The optimal configuration process of the automation terminals based on the fault observation of the whole network is shown in Fig. 6.

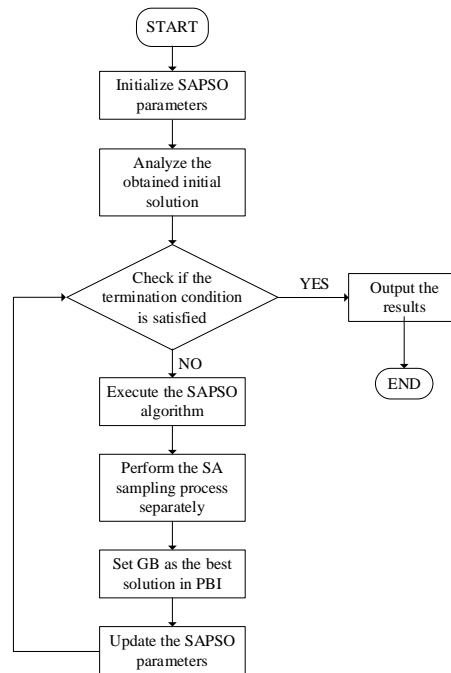


Figure 6: Flow chart of optimized configuration

Specific implementation steps are as follows:

- (1) Build a knowledge graph for the distribution network to read and assign identifiers to the attributes of nodes and lines.
- (2) According to the hierarchical rules of the distribution network, transform the topology network into a new type of layered diagram for the distribution network. Establish constraints for the fault observability of the entire network based on Eq. (1) and Eq. (2).
- (3) Calculate the line length weight based on the length of each line and Eq. (4). Simulate the historical failure rate of each line using random arrays and calculate the comprehensive weight coefficient of each node based on Eq. (5).
- (4) Construct an optimal configuration mathematical model with the goal function of minimizing the number of automation terminal configurations, combined with constraint conditions.
- (5) Transform the constrained objective function in the model into an unconstrained generalized objective function using a non-fixed multi-segment mapping penalty function.
- (6) Initialize the algorithm parameters and read in the generalized objective function for optimization and solution searching.
- (7) Output the final configuration result when the algorithm termination condition is reached.

VI. SIMULATION VERIFICATION

6.1 Layout optimization configuration of Automation terminal

This paper takes the actual 10 kV complex new distribution network as an example to optimize its automation terminal configuration, as shown in Fig. 7. Among them, E_4 and E_7 are 0.3MW wind turbines and 0.2MW photovoltaic power supplies, respectively.

- (1) B Each branch of the distribution network is numbered, and the required knowledge of it and each node is extracted under the guidance of schema layer to build the knowledge graph for the optimization of the location of the new distribution network distance ranging automation terminal. The length of each line is shown in Table 5.

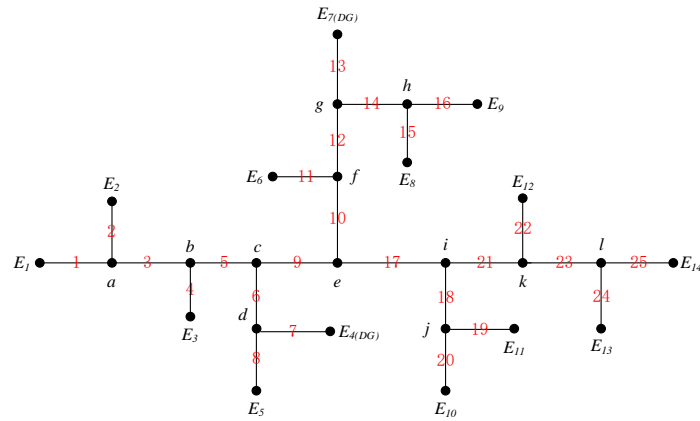


Figure 7: 10KV distribution network topology

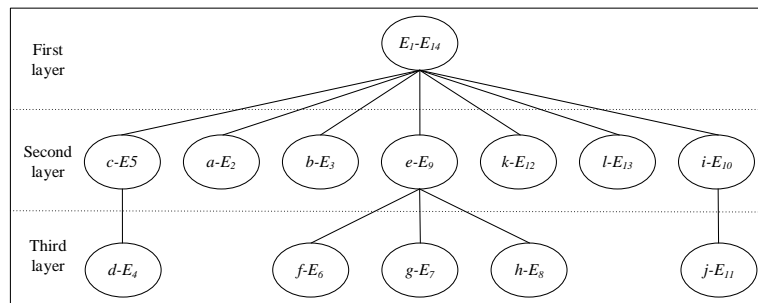


Figure 8: Stratification diagram of 10kV distribution network

Table 5: Length of line

Line	Length/km	Line	Length/km	Line	Length/km
1	1.2	2	0.8	3	1
4	1.2	5	0.8	6	1.5
7	1	8	0.6	9	1.2
10	1.8	11	1	12	1.5
13	1.2	14	1.6	15	0.6
16	0.5	17	2.5	18	1.5
19	1.2	20	1	21	1.6
22	1	23	1.8	24	1.5
25	0.8	—	—	—	—

- (2) According to the distribution network layering rules, construct the corresponding layering structure diagram in Fig. 7. Then determine the trunk line, single branch line, and multi-branch line, as shown in Fig. 8.
- (3) Calculate the weight of each line length in Fig. 7 according to Eq. (4) and line length.
- (4) Use the rand function in Matlab to generate a random number set matching the number of lines in the network to simulate the historical fault frequency of the system line, as shown in Table 6.

Table 6: Historical failure rate of each line

Line	Historical fault frequency	Line	Historical fault frequency	Line	Historical fault frequency
1	0.42	2	0.79	3	0.13
4	0.68	5	0.71	6	0.25
7	0.57	8	0.06	9	0.87
10	0.18	11	0.35	12	0.72
13	0.39	14	0.81	15	0.47
16	0.12	17	0.60	18	0.34
19	0.49	20	0.76	21	0.41
22	0.54	23	0.11	24	0.63
25	0.88	—	—	—	—

(5) Combine the weight coefficient of line fault probability and the weight coefficient of line length and obtain the weight coefficient of each line end node through Eq. (5). The results are shown in Table 7.

(6) Considering that automation terminals have been installed in E₁ and E₁₄ before the optimization configuration, the mathematical model established based on the above steps is shown in Eq. (8).

Table 7: Weights of terminal nodes of each line w_i

Node	Weight	Node	Weight	Node	Weight
E ₁	0.21	E ₆	0.51	E ₁₁	0.22
E ₂	0.25	E ₇	0.47	E ₁₂	0.49
E ₃	0.64	E ₈	0.27	E ₁₃	0.32
E ₄	0.23	E ₉	0.35	E ₁₄	0.24
E ₅	0.30	E ₁₀	0.29	—	—

$$\begin{cases} \min \sum w_i x_i, & i = 1, 2, \dots, 16 \\ x_i \in \{0, 1\}, & i = 1, 2, \dots, 16 \\ x_1 = 1, & x_{14} = 1 \\ L_j = 1, & j = 1, 2, \dots, 28 \end{cases} \quad (8)$$

(7) Set algorithm parameters, generate initial particle swarm, and perform iterative solution.

(8) The output optimization configuration result is 10010001001001. In this result, the five nodes whose value is 1 are automation terminal configuration nodes, that is, the nodes to be configured are E₁, E₄, E₈, E₁₁, and E₁₄. Since automation terminals have been configured on nodes E₁ and E₁₄, add nodes E₄, E₈, and E₁₁, as shown in Fig. 9.

According to the optimized configuration results, the number of automation terminals needed to realize the fault observation of the whole network is 5. The number of terminal nodes that do not require automation terminal has been reduced by 64%, significantly reducing investment costs. The observable situation of line faults based on this scheme is shown in Table 8. The directly observable fault lines account for 64% of the total line length of the whole network, and the indirectly observable fault lines account for 36%. The scheme can directly locate 64% of the line faults in the whole network.

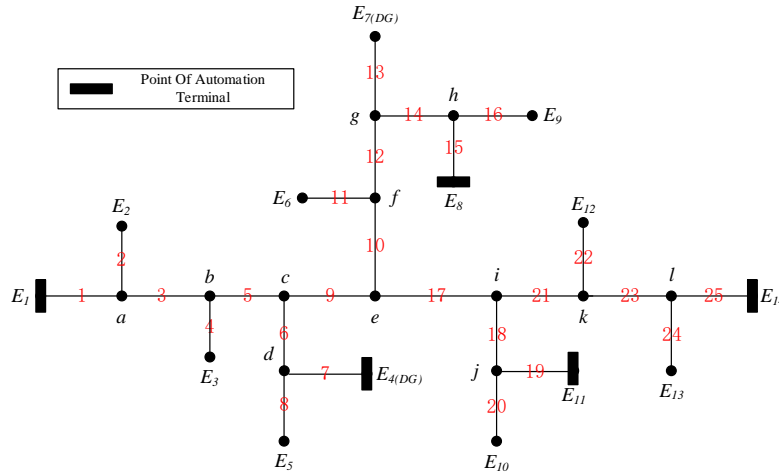


Figure 9: Optimal configuration point of automation terminal

Table 8: Observable situation of line faults

Fault directly observable line	Fault indirectly observable line
1、3、5、6、7、9、10、12、14、15、17	2、4、8、11、13、16、20、22、24
、18、19、21、23、25	

6.2 Location calculation test

PSCAD/EMTDC software was used to build a simulation model based on the distribution network, to verify the fault sensing and location capability of the distribution network under the proposed automation terminal configuration scheme, as shown in Fig. 7. The fault location is completed based on the locating method used in reference [22]. According to the optimized configuration scheme, a automation terminals is configured on five nodes: E₁, E₄, E₈, E₁₁, and E₁₄. Assume that a single-phase ground fault occurs at 600m from the indirectly observable line 13 to node E₇, as shown in Fig 10. The fault travelling wave signal was detected by 5-line automation terminals, and the time for the fault travelling wave linear mode component to reach each termianl is shown in Table 9.

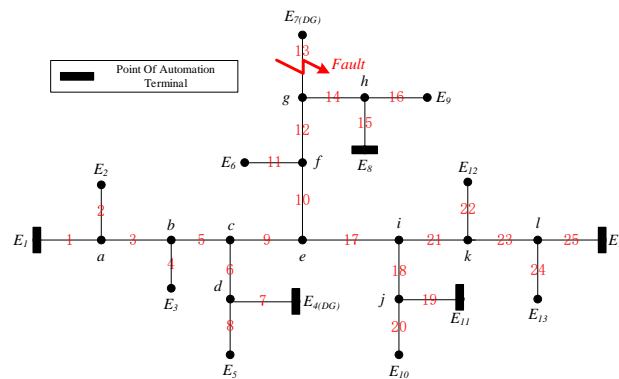


Figure 10: Simulated fault setting

Table 9: Arrival time of linear mode component

Test node	Time/ μs	Test node	Time / μs
E ₁	29.5	E ₄	28.73
E ₈	10	E ₁₁	32.5
E ₁₄	37.86	—	—

The inherent distance matrix before the fault is L obtained by each line length.

$$L = \begin{bmatrix} 0 & -3.5 & -5.3 & -7 & -9.3 \\ -3.1 & 0 & -4.8 & -6.5 & -8.8 \\ -7.3 & -7.2 & 0 & -8.3 & -10.6 \\ -7 & -6.9 & -6.3 & 0 & -5.3 \\ -8.5 & -8.4 & -7.8 & -4.5 & 0 \end{bmatrix} \tag{9}$$

The fault distance matrix d is calculated from the two-terminal location formula and the time of the fault traveling wave line mode component to each terminal.

$$d = \begin{bmatrix} 0 & -0.513 & -5.300 & 0.993 & 2.483 \\ 0.513 & 0 & -4.787 & 1.506 & 2.996 \\ 5.300 & 4.787 & 0 & 6.293 & 7.783 \\ -0.993 & -1.506 & -6.293 & 0 & 1.490 \\ -2.483 & -2.996 & -7.783 & -1.490 & 0 \end{bmatrix} \tag{10}$$

The difference between the fault distance matrix d and the inherent distance matrix L is calculated, and the branch decision matrix δ is obtained. The modified branch decision matrix obtained by correcting δ is as follows.

$$\delta' = \begin{bmatrix} 0 & 3.004 & 0 & 7.989 & 11.812 \\ 3.586 & 0 & 0 & 8.023 & 11.807 \\ 12.607 & 12 & 0 & 14.640 & 18.396 \\ 6.023 & 5.385 & 0 & 0 & 6.813 \\ 5.988 & 5.406 & 0 & 3.001 & 0 \end{bmatrix} \tag{11}$$

Since the column element corresponding to E_8 in the modified branch decision matrix is 0, it is preliminarily determined that the fault occurs in the reference node g of the branch where E_8 is located. In view of the existence of non-effective branch nodes in the distribution network, to further determine the specific location of the fault point, it is necessary to calculate the branch coefficient value W of all branches between endpoints E_4 and E_8 , as shown in Table 10.

Table 10: Branch coefficient value of each branch W

Test node	Branch	W/km
E_4, E_8	{7}	14.006
E_4, E_8	{8}	12.006
E_4, E_8	{11}	2.994
E_4, E_8	{13}	0
E_4, E_8	{14,15}	4.394
E_4, E_8	{16}	3.194

As can be seen from Table 10, the W value of branch 13 is the smallest and 0. According to the rule, the fault occurs in branch route 13. According to Eq. (12), the distance d_{E8F} between the fault point and the E_8 node can be calculated.

$$d_{E8F} = \left| \frac{(t_{E8F1} - t_{E8F0})(t_{E8F1} - t_{EnF1})}{(t_{E8F0} - t_{EnF0}) - (t_{E8F1} - t_{EnF1})} v_{E8F1} \right| \tag{12}$$

In Eq. 12: v_{E8F1} is the transmission speed of the linear mode component of the fault traveling wave, and $2.9812 \times 10^5 \text{ km/s}$ is taken in this paper. t_{E8F0} and t_{E8F1} are the transmission time of zero mode and line mode components of fault traveling wave to E_8 node respectively. t_{EnF0} and t_{EnF1} are the transmission time of zero mode and line mode components of the faulty traveling wave to E_n , respectively. In this example, E_n is E_4 . According to this calculation, the distance between the fault point and the node E_8 is d_{E8F} .

$$d_{E_8F} = \left| \frac{(10 - 10.2)(10 - 28.73)}{(10.2 - 32.85) - (10 - 28.73)} \times 2.9812 \right| = 2.849km$$

The distance from E₇ node is d_{E_7F} .

$$d_{E_7F} = L_{E_8E_7} - d_{E_8F} = 0.551km\sqrt{2}$$

The fault point is 551 m away from the E₇ node. The actual fault distance is 600 m, and the location error is 49 m, which is within the allowable range of the project. The fault of branch 13 can be accurately located.

Each indirectly observable fault line in Table 8 has the same characteristics as branch 13. The fault branches can be determined by the branch coefficient in the location method, and then the fault location can be realized by formula (12), which is not restricted by the location of the section. Therefore, the fault indirect observable line in the optimized configuration method in this paper can also be accurately located, and there is no blind spot in distance measurement.

With the development of the new distribution network, DG has been widely connected to the distribution network. To verify the adaptability of the proposed method in DG access conditions, this paper changes the type, layout, fault location, fault point and fault type of DG, and makes statistics on the simulation results in Table 11. Simulation results show that the proposed method can adapt to the change of DG access points and DG types. By updating the data layer in the knowledge graph of distribution network, the optimal location of automation terminal can be obtained flexibly. The configuration results ensure that the fault of the whole distribution network can be observed, and the fault location error is within the acceptable range.

Table 11: Statistics of simulation result

DG configuration and position	Automation terminal configuration	Fault branch	Reference node	Fault setting/m	Fault type	Location branch	Fault location/m	Error /m
E ₄ : 0.3MW wind power E ₇ : 0.2MW photovoltaic power	E ₁ 、E ₄ 、E ₈ 、 E ₁₁ 、E ₁₄	13	E ₇	600	Ag	13	521	49
		10	f	1000	AB	10	1014	14
		21	i	800	ABg	21	783	17
E ₂ : 0.3MW wind power E ₁₂ : 0.3MW battery	E ₁ 、E ₄ 、E ₈ 、 E ₁₁ 、E ₁₄	13	E ₇	600	Ag	13	540	60
		10	f	1000	AB	10	1012	12
		21	i	800	ABg	21	779	21
E ₅ : 0.2MW photovoltaic power E ₉ : 0.3MW battery	E ₁ 、E ₄ 、E ₉ 、 E ₁₁ 、E ₁₄	13	E ₇	600	Ag	13	543	57
		10	f	1000	AB	10	1015	15
		21	i	800	ABg	21	789	11

VII. CONCLUSION

An automation terminal optimal configuration method based on knowledge graph for new distribution network is proposed in this study. Through theoretical analysis, simulation and laboratory tests verification, the following conclusions may be reached.

In the construction of knowledge graph schema layer, the basic conditions of automation terminal configuration are summarized based on the fault observable definition, the least number of automation terminals is taken as the goal function, and the hierarchical analysis of distribution network is carried out. The new distribution network lines are divided into main trunk lines, single branch lines and multi-branch lines, and discussed respectively. This method is helpful to construct the constraints in the mathematical model of optimal configuration, and is not affected by new energy access.

Based on the graph computing power in knowledge graph, the optimal configuration problem is transformed into a mathematical programming problem, and a mathematical model is established with the minimum number of automation terminals as the goal function. At the same time, the model considers the present terminal configuration, line length, historical failure rate and other factors, reflecting the different terminal configuration requirements of each node. The simulated annealing particle swarm optimization algorithm is used to increase the robustness of the configuration results.

The simulation verification and laboratory verification indicate that the proposed method can assure that the faults of the complete complex distribution network can be observed, and the number of terminals required is minimum, and it can detect any fault points in the whole network. This method is practical, and it is expected to support the continued development and deployment of traveling wave locating method in the background of large-scale DG access to distribution network.

ACKNOWLEDGMENT

The authors acknowledge the support of Power Dispatching Control Center of Guizhou Power Grid Co.,Ltd

REFERENCES

- [1] Zhang, S., Liu, W., Wan, H., Bai, Y., Yang, Y., & Ma, Y., et al. (2023). Combing data-driven and model-driven methods for high proportion renewable energy distribution network reliability evaluation. *International journal of electrical power and energy systems*.
- [2] Wang H., Lu H., Sun K., Wu X W., He Y L., Du X D., (2023). Reliability evaluation method for distribution network with distributed generations considering feeder fault recovery and network reconfiguration. *IET Renewable Power Generation*, (14),3484-3495.
- [3] Islam, M. M., Usman, M. U., Newaz, A., & Faruque, M. O. (2023). Ensemble voting-based fault classification and location identification for a distribution system with microgrids using smart meter measurements. *IET Smart Grid*, 6(3), 219–232.
- [4] Leng, H., He, S., Qiu, J., Liu, F., Huang, X., & Zhu, J. (2024). Multi-Branch Fault Line Location Method Based on Time Difference Matrix Fitting. *Energy Engineering*, 121(1), 77–94.
- [5] Azeroual, M., Boujoudar, Y., Aljarbouh, A., Moussaoui, H. E., & Markhi, H. E. (2022). A multi-agent-based for fault location in distribution networks with wind power generator. *Wind Engineering*, (3), 46.
- [6] Ananthan, S. N., & Santoso, S. (2020). Application technique for model-based approach to estimate fault location. *IET Smart Grid*, (4), 3.
- [7] Chen S. Y., Feng Y., Zheng Z. X., Xu Y. F., Cui J. Y., Le J. (2022) Optimal configuration method of fault location monitoring device in low voltage active distribution network based on node participation. *Electrical Measurement & Instrumentation*.
- [8] Xia, S., Wang, Z., Gao, X., & Li, W. (2024). Optimal planning of mobile energy storage in active distribution network. *IET Smart Grid*, 7(1), 1–12.
- [9] Zhang Q., Li W. J., Zhang Z. H., Li Y. (2021). Optimal Configuration for Fault Indicator of Distribution Network Based on PSO Algorithm. *Power Capacitor & Reactive Power Compensation*, (6), 42.
- [10] Zhou M., Mo J., Yao Y., Chen Y. B., Wang X. D., Zhao X. L. (2019). Research on FTU optimal configuration model of distribution network based on fault observability index. *Power System Protection and Control*, (17), 47.
- [11] Yuan Y. L., Li L. (2021). Optimized Configuration Model for the Fault Locating System of Charging Piles for Electric Vehicles. *Electric Vehicle Technology*, (2), 43.
- [12] Liu L., Yan X. M., Li G. (2023). Knowledge Graph in Distribution Network Fault Handling: Advances, Challenges and Prospects. *Electric Power Information and Communication Technology*, (7), 21.
- [13] Song W. Q., Han L., Xian H. Z., Yao S. N., Guo S. (2022). Construction and application of GCN-based distribution network knowledge graph. *Electronic Design Engineering*, (7), 30.
- [14] Lyu X. D., Gao H. J., Ye S. Y., Li H. B., Yang R., Liu J. Y., Sun H. J. (2022). Intelligent Terminal Planning Strategy Considering Reliability and Economy Reconfiguration for Distribution Network. *Proceedings of the CSEE*, (2), 42.
- [15] Bian X. Y., Zhang L. Y., Zhou B., Xu B., Lin S. F. (2022). Review on Domestic and International Electricity Market Research Based on Knowledge Graph. *Transactions of China Electrotechnical Society*, (11), 37.
- [16] Li Y., Li C., Wang Y., Guo X. X., Li L., Ma J. Y. (2023). Optical Cable Fault Diagnosis and Auxiliary Decision-making Based on Knowledge Graph. *Journal of Physics: Conference Series*, (1), D.
- [17] Li Z. B., Huang Z. G., Guo L. X., Shan L. F., Yu G. Y., Chong Z. Q., Zhang Y. (2023). Cognitive knowledge graph generation for grid fault handling based on attention mechanism combined with multi-modal factor fusion. *Computers and Electrical Engineering*, (PA).

- [18] Pu T. J., Tan Y. P., Peng G. Z., Xu H. F., Zhang Z. H., (2021). Construction and Application of Knowledge Graph in the Electric Power Field. *Power System Technology*, (6), 45.
- [19] Gao Z. P., Zhao Y., Yu Y.L., Luo Y. J., Xu Z. W., Zhang L. M., (2020). Knowledge graph-based method for identifying topological structure of low-voltage distribution network. *The Journal of Engineering*, (2), 48.
- [20] Wang J. D., Yang J., Pei Y. Z., Zhan X. P., Zhou T., Xie P. Y. (2021). Distribution Network Fault Assistant Decision-making Based on Knowledge Graph. *Power System Technology*, (6), 45.
- [21] Guo R., Yang Q., Liu S. H., Li W., Yuan X., Huang X. H. (2021). Construction and Application of Power Grid Fault Handling Knowledge Graph. *Power System Technology*, (6), 45.
- [22] Xie L. W., Li Y., Luo L. F., Chen C., Cao Y. J. (2020). Fault Location Method for Distribution Networks Based on Distance Matrix and Branch Coefficient. *Proceedings of the CSEE*, (7), 40.
- [23] Cai C., Ding J. Z., Lyu F., Yuan H. X., Zhu Y., Sun G. Q., Wei Z. N. (2020). Optimal placement of fault indicators based on integer linear programming model in distribution network. *Power System Protection and Control*, (1),48.
- [24] Liu, M. (2022). A Value-at-Risk Based Approach for PMU Placement in Distribution Systems. *Energy Engineering*, 119(2), 781–800.