
Research article

Construction and application of distribution network fault disposition knowledge graph based on multi-source data

Xingyu Wang¹ and Hongpeng Liu^{2,*}

¹ Northeast Electric Power University, 169 Changchun Road, Chuanying District, Jilin City, Jilin Province, 132012, China

² Department of Electrical Engineering, Northeast Electric Power University, 169 Changchun Road, Chuanying District, Jilin City, Jilin Province, 132012, China

* **Correspondence:** Email: liuhp_neepu@163.com; Tel: +86-15143216966.

Abstract: As digital grid development in Guangdong Province advances, the accumulation of multi-source heterogeneous data and the low efficiency of fault disposition in distribution networks have become increasingly prominent issues. To facilitate the effective use of multi-source heterogeneous data and efficient decision-making for distribution network fault disposition, this paper proposes a method for constructing a fault disposition knowledge graph based on multi-source heterogeneous data. First, the characteristics of multi-source heterogeneous data in distribution networks were analyzed, and a framework for constructing a fault disposition knowledge graph was designed accordingly. Then, named entities were extracted from unstructured fault texts using EBERT-BiLSTM-CRF. Based on the recognized entities, BiLSTM-GA was used to identify semantic relations and generate knowledge triples. Finally, fault data from a distribution network in a specific region of Guangdong Province were used in a case study, and the resulting knowledge graph was visualized and managed using the GridGraph graph database. Within the proposed knowledge graph construction framework, the complete BiLSTM-CRF-GA extraction configuration achieved a mean precision of 91.16%, a mean recall of 92.94%, and a mean F1-score of 92.06% across five matched data splits, supporting its effectiveness in extracting fault-disposition knowledge from multi-source data.

Keywords: distribution network fault disposition; knowledge graph; multi-source heterogeneous data; knowledge extraction; neural network.

1. Introduction

The ongoing digital transformation of distribution networks (DNs) has led to a substantial expansion of data resources in distribution systems, encompassing temporal, spatial, and diverse data types [1]. Concurrently, the increasing number of system devices, driven by the growing integration of renewable energy and the expansion of DNs, has resulted in a corresponding increase in fault occurrences [2]. As a critical link between the upstream grid and end-user demand, the reliable and secure operation of the DN is paramount [3]. Therefore, constructing a knowledge graph (KG) for DN fault disposition based on multi-source data has significant research value. In this paper, fault disposition refers to the complete operational process of identifying and analyzing a distribution network fault, determining its cause, formulating and implementing the corresponding corrective measures, and evaluating the post-fault operating state.

1.1. Literature review

Knowledge graph (KG) technology is an effective approach for describing complex relationships among concepts and entities in the real world. It has attracted widespread attention because of its powerful knowledge representation and reasoning capabilities [4]. Recent review studies have systematically summarized advances in KG construction, knowledge representation, and reasoning techniques, highlighting the importance of integrating heterogeneous information sources and improving intelligent knowledge utilization capabilities [5,6]. Existing KG-based fault diagnosis studies have achieved promising results in knowledge organization, fault identification, and reasoning assistance. Reference [7] developed a fault diagnosis KG construction method based on distantly supervised relation extraction, which alleviates the influence of noisy multi-source corpora and improves relation extraction performance under limited data conditions. However, the method mainly focuses on extracting entity relationships from fault-related texts and does not further consider semantic differences among heterogeneous operational data sources.

Recent data-driven fault diagnosis studies have combined hybrid nonlinear–linear feature modeling with an improved graph convolutional network for fault detection and diagnosis [8], integrating class-discriminative and spatial feature extraction with an improved graph convolutional network for fault classification [9]. However, these approaches mainly learn discriminative patterns from operational data and do not explicitly organize fault entities, causal relationships, and disposal measures for knowledge-based reasoning. Reference [10] constructed a multi-node KG using a hybrid strategy that combines data-driven and knowledge-driven approaches. By integrating domain knowledge with data analysis methods, the proposed approach improves fault detection performance. Nevertheless, this type of method still relies heavily on predefined domain knowledge and structured information, which may limit its adaptability when processing continuously generated, heterogeneous fault disposition data. Reference [11] introduced graph convolutional networks into KG-based fault diagnosis and utilized historical maintenance data to extract structural and semantic information. The method improves fault-cause tracing through multi-hop reasoning. However, GNN-based approaches mainly focus on graph representation learning and reasoning based on existing graph structures, while the automatic extraction and integration of knowledge from heterogeneous raw data remain insufficiently addressed. Reference [12] constructed a KG by extracting multi-domain features from bearing vibration signals and established a data accumulation strategy to expand the training dataset.

Although this method demonstrates the potential of combining heterogeneous information with KGs, it mainly focuses on equipment condition-monitoring scenarios rather than complex fault disposition processes in DNs.

In the context of power system fault disposition, reference [13] modeled power-event elements and their relationships using an ontology and extended KG representation by incorporating temporal, spatial, and causal information. This method enhances the expressiveness of power fault knowledge. However, ontology-based approaches require expert-defined knowledge structures and may encounter difficulties in large-scale heterogeneous data integration. Reference [14] proposed a joint entity-relation extraction method based on pretrained models for grid dispatch fault disposition and constructs a decision-support KG. Although this method improves the automation of knowledge extraction, the input data are mainly treated as homogeneous textual information, without fully considering semantic heterogeneity among different operational data sources. Reference [15] introduced multimodal constraints into power-equipment representation learning and constructed a cognitive KG for grid fault disposition by combining attention mechanisms with knowledge representation models. This method demonstrates the advantages of multimodal information integration. However, existing multimodal KG approaches mainly emphasize feature fusion and representation learning, while the semantic organization and knowledge extraction requirements of different data sources in DN fault disposition scenarios remain insufficiently considered. Reference [16] established a knowledge organization framework by defining concepts and relationships among entities, providing support for fault information analysis and intelligent decision-making. Nevertheless, this method mainly focuses on knowledge modeling and organization, while the automatic extraction of fault knowledge from heterogeneous data sources requires further improvement.

The studies above propose several constructive approaches and have achieved promising results in constructing KGs for DN fault disposition; however, several issues in this field still require attention. First, although DN fault data span multiple dimensions and are available in substantial volumes, existing studies often fail to adequately account for the effects of multi-source heterogeneity when developing fault disposition KGs. Second, the knowledge extraction capabilities of existing DN fault disposition KGs remain insufficient in terms of accuracy and precision, thereby limiting their reliability and practical usefulness in fault-disposition decision support. Third, existing BERT/ERNIE-BiLSTM-CRF-based knowledge graph construction methods generally treat input texts as homogeneous semantic sequences and mainly focus on entity recognition or relation extraction in specific domains. They rarely consider semantic heterogeneity and differences in knowledge granularity across multiple data sources. In distribution network fault disposition scenarios, equipment ledgers, fault records, dispatch protocols, and defect databases contain different types of operational knowledge. Therefore, directly applying existing methods may result in incomplete knowledge representation and insufficient fault reasoning capability.

1.2. Contributions of this work

To address the limitations noted above, this paper presents a methodology for developing KGs for DN fault disposition using multi-source data. The main contributions are as follows:

(1) A multi-source heterogeneous knowledge extraction framework is developed for DN fault disposition that considers the semantic characteristics of different data sources and achieves a unified representation of fault-related knowledge from heterogeneous information.

(2) Unlike existing BERT/ERNIE-BiLSTM-CRF-based knowledge graph methods that mainly focus on homogeneous text processing or single-source information extraction, this study develops a multi-source-oriented knowledge extraction framework for distribution network fault disposition. The proposed framework accounts for semantic differences among equipment ledger data, fault disposition records, dispatch protocols, and defect information, and integrates these heterogeneous characteristics into the knowledge graph construction process.

(3) To enhance the practical applicability of the constructed KG, a knowledge-driven fault disposition application framework is developed based on the GridGraph graph database, enabling semantic information retrieval and auxiliary fault diagnosis by exploiting the relationships among fault entities, fault causes, and corresponding corrective and disposal measures.

2. Multi-source data for DN

With the advancement of DN condition-monitoring technology and the increasing complexity of network structures, the data collected by intelligent power equipment are characterized by large volumes, rapid updates, multi-source heterogeneity, and low-value density. For example, considerable quantities of diverse textual data concerning distribution network operation are stored in systems such as the DN production management system, the data acquisition and monitoring control system, and the distribution management system [17].

This study uses multiple data sources, including the equipment ledger, fault disposition records, dispatch operation logs, and DN defect reports, to build a KG for fault disposition. The datasets are listed in Table 1.

Table 1. Multi-source data for DN.

Data name	Data type	Optional part
DN equipment ledger data	Semi-structured	Name of equipment, station, maintenance team, person in charge of the equipment, years of use, etc.
Fault disposition data	Unstructured	Fault line, cause of fault, type of fault, outage area, fault condition, etc.
Dispatch protocol data	Unstructured	New energy station grid integration management, large user grid integration management, operation, accident handling, dispatching rules, new equipment commissioning management, etc.
DN defects database	Semi-structured	Equipment name, responsible team, defect level, defect content, etc.

(1) Semi-structured data processing: The native formats of both the DN equipment ledger and defect archive conform to the structural requirements of knowledge-triplet construction. Therefore, the equipment name and defect content are taken as entity objects, while the remaining information is recorded as additional attributes.

(2) Unstructured fault data processing: Records describing fault conditions, handling processes, and causes in the fault disposition data consist of manually recorded text. They are processed using the EBERT-BiLSTM-CRF model for named entity recognition and organized into knowledge triples after processing.

(3) Dispatch protocol data processing: These data consist of rule-based unstructured text. Based on their writing logic, a “keyword + short text” approach is used to construct the KG.

3. KG construction

This study constructs a fault-disposition KG from multi-source DN data through the procedure shown in Figure 1. The schema layer is first defined from the structure of planning documents and domain-expert knowledge. The resulting schema then guides the bottom-up development of the data layer through knowledge extraction and fusion [18]. The main contribution is the knowledge extraction method for heterogeneous DN fault-disposition data. GridGraph is used as the storage and application platform for knowledge retrieval and fault diagnosis, rather than as a methodological innovation.

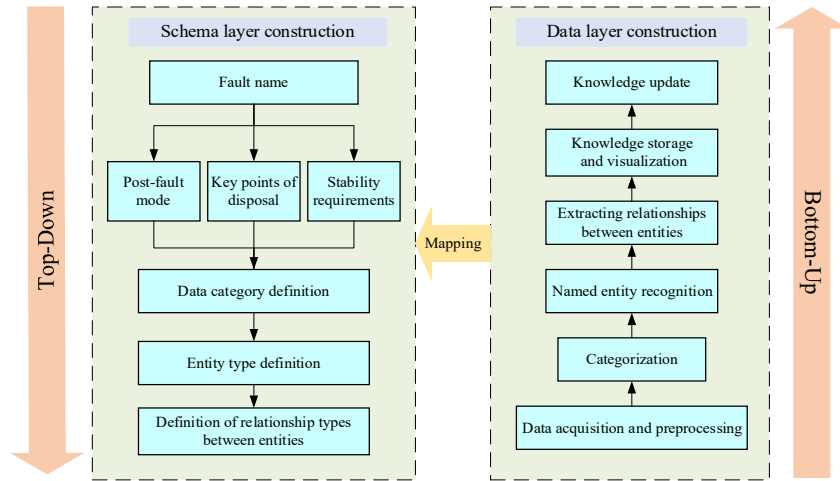


Figure 1. Top-down and bottom-up construction process of the KG for fault handling in DN.

3.1. Schema layer construction

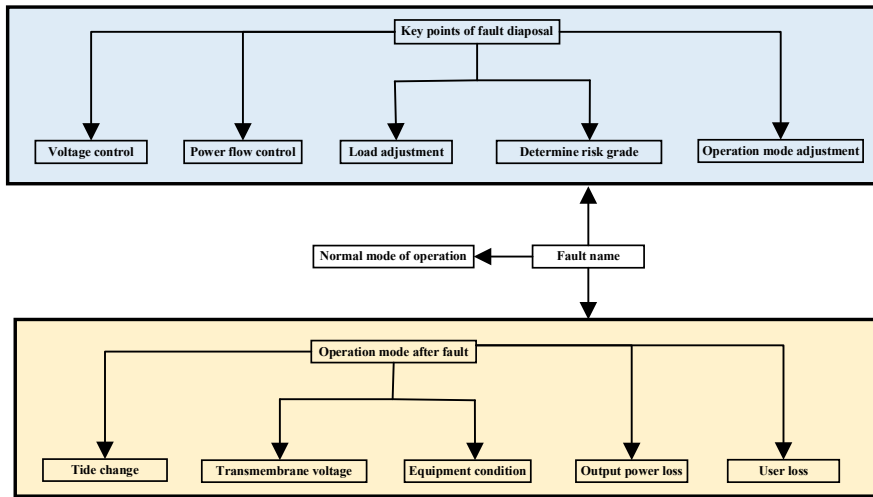


Figure 2. Pattern layer architecture of the KG for DN fault disposition.

The model layer defines the entities, relationships, and attributes required by the grid fault-disposition KG [19]. As shown in Figure 2, it represents the organizational structure of the planning text through four components: fault designation, standard operating mode, post-failure mode, and essential corrective actions, together with the relationships among them. Power-grid expertise is

used to identify concepts and attributes with practical engineering value, thereby establishing the knowledge framework for power-system faults and guiding data-layer development. The data layer generated through knowledge extraction and related procedures is then mapped to the corresponding concepts in the concept layer, allowing the knowledge structure to be updated as new information is incorporated.

3.2. Data layer construction

Existing gated attention mechanisms mainly focus on selecting globally important textual features based on hidden representations. However, DN fault disposition texts usually contain multiple entities and complex relationships, in which different relations require different contextual information. Therefore, the gated attention mechanism adopted in this study is designed specifically for entity-relation extraction tasks and aims to enhance the representation of relationship-related semantic features while reducing interference from irrelevant information.

Building the data layer involves three primary stages: knowledge extraction, knowledge integration, and graph updating. Specifically, knowledge extraction involves identifying named entities from unstructured texts using an entity recognition model aligned with the schema layer and then structuring the output; knowledge fusion centers on two fundamental tasks, namely entity linking and knowledge merging, to eliminate data conflicts and realize cross-source data correlation; and knowledge updating involves refining and adjusting the graph in response to evolving information and the continuous inflow of new content after the initial graph is constructed.

This work presents a two-stage modeling approach for power knowledge extraction. The first stage uses an EBERT-BiLSTM-CRF named entity recognition model as the foundation to accurately extract entities from small-scale power-domain data. In the second stage, an entity-relation identification model enhanced with a gated attention mechanism is introduced to extract entity relations from DN fault-disposition plans. EBERT serves as the contextual encoder in the entity recognition stage. By contrast, the main contribution of this study lies in the multi-source knowledge extraction framework and the relation-oriented gated attention mechanism.

To provide a task-specific mathematical formulation of the proposed two-stage knowledge extraction process, let $(X = (x_1, x_2, \dots, x_n))$ denote a DN fault-handling text and let $(E = e_1, e_2, \dots, e_m)$ denote the entity set identified by the EBERT-BiLSTM-CRF model. The named entity recognition stage can be expressed as

$$E = f_{\text{NER}}(X) \quad (1)$$

Based on the identified entities, the relation extraction stage determines the semantic relation r_{ij} between entities e_i and e_j . The resulting KG triple set is defined as

$$\mathcal{T} = \{(e_i, r_{ij}, e_j) \mid e_i, e_j \in E, r_{ij} = f_{\text{RE}}(X, e_i, e_j)\} \quad (2)$$

where f_{NER} represents the EBERT-BiLSTM-CRF-based named entity recognition process, f_{RE} represents the gated attention-based relation extraction process, and $\mathcal{T}$ denotes the set of triples used to construct the KG. This formulation connects entity recognition and relation extraction within a unified task-oriented framework for DN fault handling.

The target datasets encompass equipment ledger records, fault disposition plans, dispatch protocols, and defect-library entries from the DN. Specific extraction procedures are described as follows:

(1) Knowledge extraction from DN equipment ledger data.

This paper focuses on the semi-structured and highly structured characteristics of DN equipment ledger data. After the textual content is preprocessed, customized linking rules are designed according to the schema layer and the inherent data format. These rules are then used to build the KG. Representative examples of specific entities are shown in Table 2.

Table 2. Example of a DN equipment entity.

Type of entity	Entity examples
Bus route	110 kV Duan zhou A line
Generating line	Antelope Hill Station Bus
Distribution transformers	4 boxes and 3 transformers
Interrupter	231 circuit breakers
Voltage transformer	Voltage transformers on busbar side
Switchgear	10 kV inlet switchgear

(2) Knowledge extraction from fault disposition data.

While general-purpose bidirectional encoder representations from transformers (BERT) provides robust contextual representations, its general-domain pretraining corpus contains limited terminology specific to distribution networks. Consequently, specialized expressions concerning equipment states, fault causes, and fault disposition measures may not be adequately represented. To alleviate this domain mismatch, this study employs enhanced BERT (EBERT), a BERT-based encoder adapted to distribution-network fault texts. EBERT provides domain-oriented contextual representations for subsequent named entity recognition. In the proposed architecture, EBERT is used as the contextual encoding component rather than being presented as a newly developed standalone model.

In this study, the EBERT-BiLSTM-CRF model is used to identify named entities in unstructured fault disposition data. Through structural analysis of the fault disposition information, entities such as location, voltage level, involved equipment, fault cause, treatment process, and operating status after treatment are identified. Representative entities are listed in Table 3.

Table 3. Examples of fault disposition entities.

Type of entity	Entity examples
Fault location	XXX line, XXX switching station
Voltage level	10 kV
Faulty equipment	Intelligent ring maintainer, box transformer, circuit breaker, ice melting device
Fault cause	Cable insulation deterioration, bird short circuit, equipment dirt flashover
Fault disposition measure	Replace the fault indicator, adjust the protection value, remove foreign matter
Operating state after fault disposition	Load optimization for Taiwan, ring power restoration

(3) Knowledge extraction from dispatch protocol data.

Although the dispatch protocol data examined here are essentially unstructured, they follow a concise and coherent format, with most records appearing as brief textual entries. Accordingly, a

manual “keyword + short text” approach is used to construct the KG of scheduling regulations. Representative examples are shown in Table 4.

Table 4. Examples of scheduling protocol entities.

Type of entity	Entity examples
System administration tasks	Grid connection management, incident handling, load dispatch management, maintenance management
Switch operation	The relay protection must be fully operational before the switch is energized
Line operation	Line outage operation disconnect switch, pull line side knife gate
Troubleshooting measures for unusual fluctuation indicators	Start the standby unit

The EBERT-BiLSTM-CRF model consists of three components [20].

- The EBERT pretrained language model primarily uses EBERT’s tokenizer and encoder to segment and encode the input text. Specifically, the input is an utterance, and the output consists of the corresponding word vectors, which serve as inputs to subsequent models.
- A BiLSTM employs two long short-term memory (LSTM) layers that traverse the input sequence in opposite directions. Their outputs are combined to produce a contextual representation containing information from both directions. Compared with conventional recurrent neural networks (RNNs), this structure is less susceptible to vanishing and exploding gradients when modeling dependencies across long sequences.

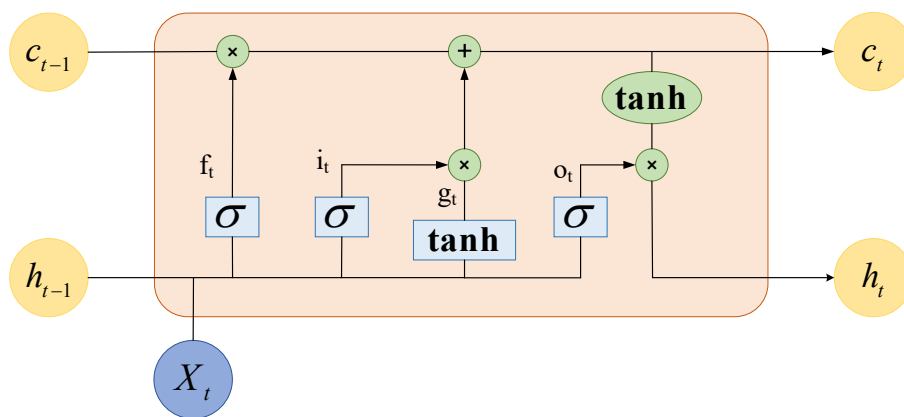


Figure 3. LSTM hidden cell.

Figure 3 presents the sequence of computations within an LSTM cell. The forget gate is evaluated first using the current input and the output carried forward from the previous time step. Its formulation is given below:

$$b_f f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad (3)$$

where t denotes the time step; f_t , x_t , and h_{t-1} correspond to the forget gate, the current input, and the hidden state from the preceding time step, respectively. W_f and b_f are the trainable weight

matrix and bias vector of the forget gate, respectively; and $\sigma(\cdot)$ denotes the sigmoid activation function. The notation $[h_{t-1}, x_t]$ denotes the concatenation of h_{t-1} and x_t .

The input gate processes both the current input and the output from the preceding unit, updating the internal state to facilitate a new cycle of information capture. The corresponding model is described below:

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (4)$$

where i_t denotes the input-gate vector, while W_i and b_i are the corresponding trainable weight matrix and bias vector, respectively.

$$\tilde{C}_t = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \quad (5)$$

where $\tilde{C}_t$ denotes the candidate cell-state vector; W_c and b_c are the corresponding trainable weight matrix and bias vector, respectively; and $\tanh(\cdot)$ denotes the hyperbolic tangent activation function.

$$C_t = f_t \cdot C_{t-1} + i_t \cdot \tilde{C}_t \quad (6)$$

where C_t and C_{t-1} denote the current and previous cell-state vectors, respectively, and multiplication between a gate vector and a cell-state vector denotes element-wise multiplication.

Finally, the output gate is computed using the current input and the preceding cell's output, as shown in the following formulation:

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \quad (7)$$

$$h_t = o_t * \tanh(C_t) \quad (8)$$

where o_t denotes the output-gate vector, while W_o and b_o are the corresponding trainable weight matrix and bias vector, respectively; h_t denotes the hidden-state vector at time step t , and the asterisk denotes element-wise multiplication.

A conventional LSTM processes information only in the forward direction. However, contextual information in DN fault texts is correlated in both directions. Therefore, BiLSTM introduces an independent reverse LSTM to obtain bidirectional textual semantic representations. The BiLSTM architecture is depicted in Figure 4, with its detailed configuration presented below:

$$H_t[h_t^{fw}; h_t^{bw}] \quad (9)$$

where h_t^{fw} denotes the hidden state of the forward LSTM at time step t ; h_t^{bw} is the hidden state of the reverse LSTM at time step t ; H_t is obtained by merging the two directional hidden states, and the resulting vector sequence is expressed as follows:

$$H_t = \{h_0, h_1, \dots, h_n\} \quad (10)$$

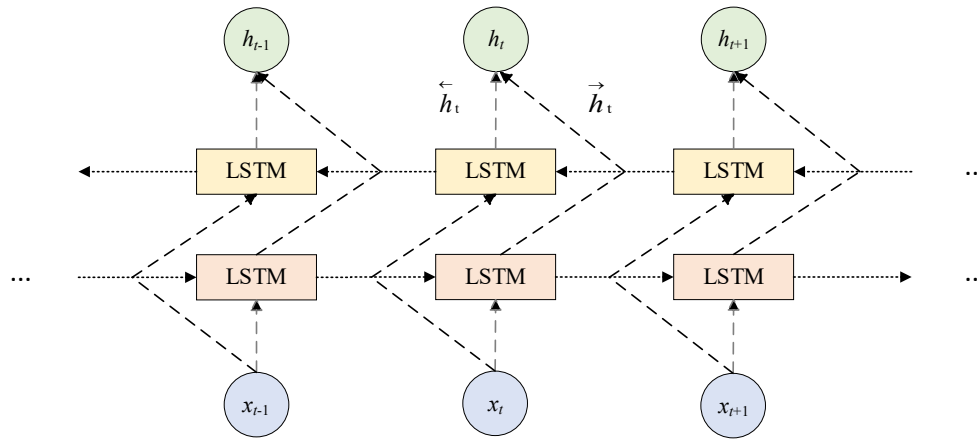


Figure 4. The BiLSTM model architecture.

(c) CRF modeling layer. The CRF model is used to obtain the optimal annotation sequence at the utterance level and output the corresponding named-entity labels. The conditional probability is defined as follows [21]:

$$p(y|x) = \frac{1}{Z} \exp\left(\sum_j \sum_{i=1}^{n-1} \lambda_j t_j(y_{i-1}, y_i, x, i) + \sum_k \sum_{i=1}^{n-1} u_k s_k(y_i, x, i)\right) \quad (11)$$

where x denotes the observed input sequence; y denotes a candidate label sequence; y_i and y_{i-1} are the labels at positions i and $i-1$, respectively; and $p(y|x)$ denotes the conditional probability of label sequence y given input sequence x . The variable i indexes the position in the sequence, while j and k index the transition and state feature functions, respectively. $s_k(y_i, x, i)$ denotes the k -th state feature function. Here, $t_j(y_{i-1}, y_i, x, i)$ is the transition feature function between neighboring label sequences in the observed sequence; λ_j and u_k are the weights corresponding to the transition and state feature functions, respectively; n is the length of the sequence; and Z is the normalization factor used to ensure that the conditional probability is correctly defined, as expressed below:

$$Z = \exp \sum_y \left(\sum_j \sum_{i=1}^{n-1} \lambda_j t_j(y_{i-1}, y_i, x, i) + \sum_k \sum_{i=1}^{n-1} u_k s_k(y_i, x, i) \right) \quad (12)$$

In the complete two-stage knowledge extraction configuration, the CRF layer is used only to decode the entity-label sequence in the named entity recognition stage and does not directly predict relation types. Relation prediction is performed in the subsequent BiLSTM-GA component, where gated attention extracts relation-relevant features, and a sigmoid output layer predicts the corresponding relation types. Attention mechanisms optimize model performance by assigning different weights to sequence elements according to their importance. Traditional attention networks assign weights to every element in a sequence, even when some elements are less relevant to the task, which can distract the model from key information. This limitation can cause a marked decline in model performance, especially when processing lengthy sequences. To address this issue, this study integrates a gating mechanism into the attention framework. The gating unit filters relevant elements, calculates weights only for selected elements, and automatically ignores irrelevant information. As shown in Figure 5, the gated attention architecture comprises two main components: a backbone network and a supporting auxiliary network.

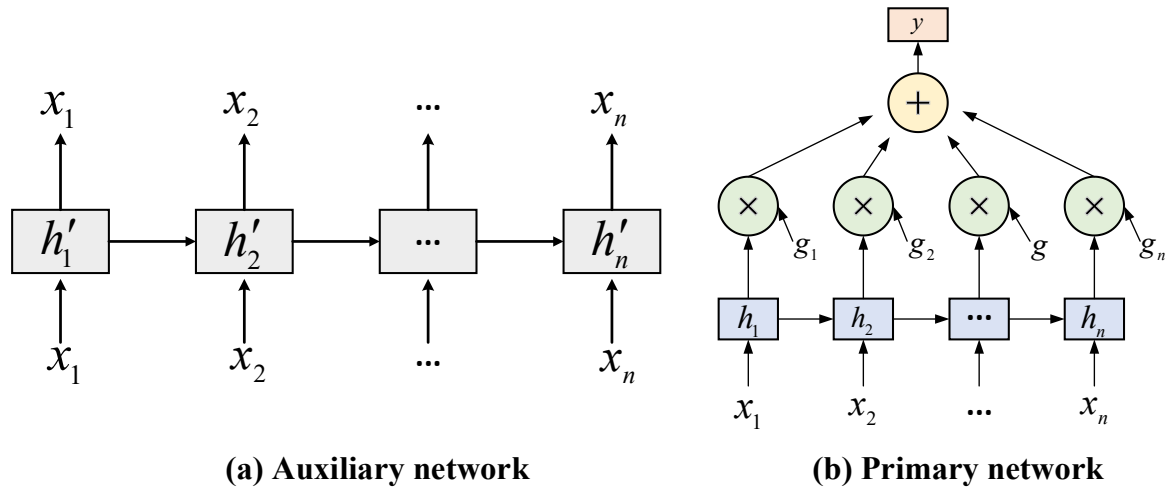


Figure 5. Auxiliary and backbone networks of the gated attention mechanism.

The network on the right is the backbone network. The input sentence sequence $x = [x_1, x_2, \dots, x_n]$ is processed by the BiLSTM to obtain $h = [h_1, h_2, \dots, h_n]$. A gating mechanism $G = \{g_1, g_2, \dots, g_n\}$ is then introduced, where $g_n \in \{0, 1\}$ is associated with each time step. When g_t equals 1 at time step t , the gate is open; when it equals 0, the gate is closed. This controls whether information from the current state is considered. Then, for the set of positions S corresponding to a sequence of $g_t = 1$, the attention score can be computed as follows:

$$e_t = \text{MLP}(h_t) \quad t \in S \quad (13)$$

where a multilayer perceptron (MLP) is used to learn a nonlinear combination of features, and the attention weights are calculated using Eq (14), which incorporates a gating mechanism into the attention computation.

$$\alpha_t = \frac{\exp(e_t)}{\sum_{t' \in S} \exp(e_{t'})}, \quad \sum_{t \in S} \alpha_t = 1 \quad (14)$$

where α_t denotes the normalized attention weight assigned to position t , and t' is a dummy index that runs over all selected positions in S . The auxiliary network on the left generates these binary gates. As shown in Figure 6, it uses the same input features as the backbone model but adopts a more compact network architecture. The auxiliary network produces a series of probability $p = \{p_1, p_2, \dots, p_n\}$ values, calculated as follows:

$$h'_t = \text{BiLSTM}(h'_{t-1}, x_t) \quad (15)$$

$$p_t = \text{sigmoid}(Uh'_t) \quad (16)$$

where h'_t and h'_{t-1} denote the current and previous hidden states of the auxiliary BiLSTM network, respectively. The prime symbol distinguishes the hidden states of the auxiliary network from those of the backbone network. Here, U is the weight matrix, and the sigmoid function maps the result to a value between 0 and 1; p_t determines the probability that the gate is open. The binary gate is sampled from a Bernoulli distribution, as shown in the following equation:

$$g_t \leftarrow \text{Bernoulli}(p_t) \quad (17)$$

After the binary gates and gated attention weights are obtained, the contextual relation representation is calculated by aggregating the BiLSTM hidden states as follows:

$$\mathbf{c} = \sum_{t=1}^n \alpha_t \mathbf{h}_t \quad (18)$$

where $\mathbf{h}_t$ denotes the contextual feature generated by the BiLSTM at time step t , and α_t denotes the normalized attention weight defined in Eq (14) over the positions selected by the binary gates in Eq (17). Therefore, hidden states corresponding to closed gates do not contribute to the relation representation. The resulting representation is then mapped to the predicted relation types:

$$\hat{\mathbf{r}} = \text{sigmoid}(\mathbf{W}_R \mathbf{c} + \mathbf{b}_R) \quad (19)$$

where $\mathbf{W}_R$ and $\mathbf{b}_R$ are the trainable parameters of the output layer, and $\hat{\mathbf{r}}$ represents the predicted probabilities of the relation types. In this manner, the gated attention mechanism connects the dynamic selection of textual features with the final relation extraction results.

Using the auxiliary and backbone networks, the GA dynamically selects a subset of elements for computation and calculates attention weights for the selected elements, enabling effective information aggregation. For DN fault texts, this mechanism ensures that the model's attention focuses on key information, such as faulty equipment or risk points, thereby reducing interference from non-core elements. Combined with BiLSTM's ability to capture long-term dependencies, the model can identify causal links across multiple sentences in fault texts, providing reliable support for relationship extraction.

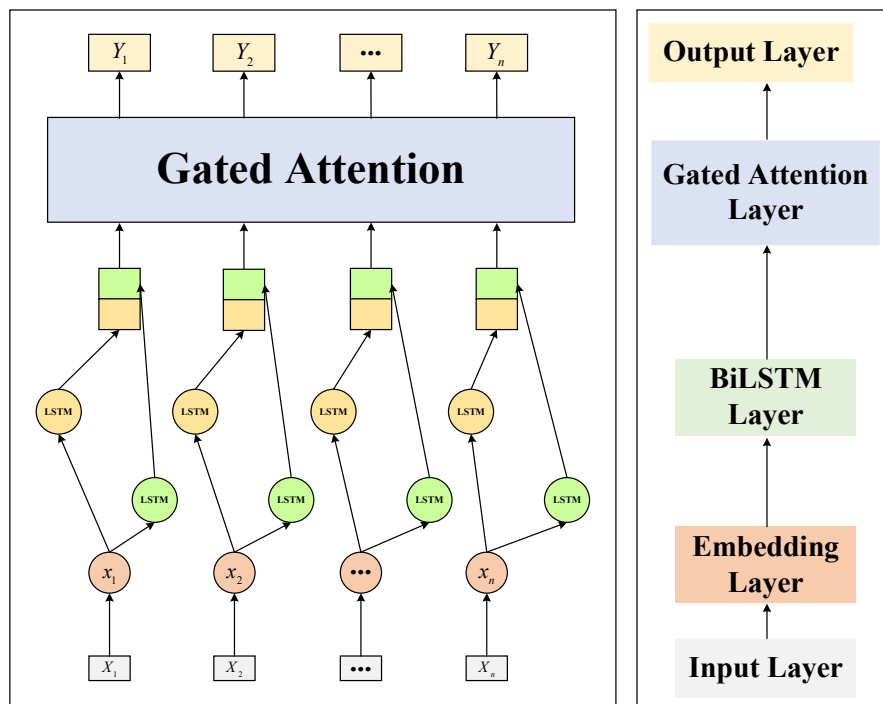


Figure 6. BiLSTM-GA model structure.

Figure 6 illustrates the architecture of the BiLSTM-GA relation extraction component. In the complete two-stage knowledge extraction workflow, the EBERT-BiLSTM-CRF component first identifies named entities and their corresponding label sequences. Based on the identified entities, BiLSTM-GA extracts semantic relations between entity pairs. The model consists of five hierarchical components, beginning with the input layer, which receives the entity-related fault text. The embedding layer first maps the entity-related fault text to word-level vectors. These vectors are then passed to the BiLSTM layer, which reads the sequence in both directions so that information from distant positions can contribute to its contextual representation. The output layer subsequently applies a sigmoid activation function to predict the corresponding distribution-network fault relation types. In the comparative experiments reported below, the complete extraction configuration is denoted as BiLSTM-CRF-GA, where CRF refers to the upstream entity-sequence decoding component, and GA refers to the downstream relation extraction component.

In summary, the proposed framework does not modify the basic internal formulations of LSTM, BiLSTM, or CRF. Its task-specific contribution lies in integrating contextual entity representations, sequence-level entity decoding, and gated feature aggregation into a two-stage process for extracting KG triples from DN fault-handling texts. The EBERT-BiLSTM-CRF model identifies domain entities, while the gated attention mechanism reduces the influence of non-critical textual information during relation extraction. The entity and relation outputs are subsequently combined to generate structured triples for KG construction.

Knowledge fusion eliminates redundant or inaccurate information by aligning two KGs through entity disambiguation and linking, thereby producing a unified knowledge representation. Knowledge updating, by contrast, begins by processing preplan texts to obtain an ontology. Subsequently, predefined reasoning rules are used to infer new relationships from known named entities and their relations, followed by a final evaluation of the reasoning outcomes [22].

4. Example analysis

4.1. Parameter description

The dataset was obtained from practical distribution network fault-handling scenarios. It includes equipment ledger records, fault-handling reports, dispatching protocols, and defect information. After preprocessing and expert screening, 496 fault records comprising 976 text samples were retained for entity recognition and entity-relation extraction tasks.

Although the dataset is relatively small compared with general natural language processing datasets, DN fault-handling data exhibit strong domain specificity and specialized characteristics. Each sample contains detailed information related to fault types, equipment states, fault causes, and disposal measures. Therefore, the complexity of knowledge extraction depends not only on the number of samples but also on the diversity of domain-specific semantic relationships.

The text data were annotated at the character level using the begin-inside-outside (BIO) tagging scheme and were subsequently divided into training, validation, and testing sets at a ratio of 8:1:1. Due to confidentiality requirements and data-security regulations governing power system operation, the original fault records cannot be publicly released. However, the data sources, data composition, annotation strategy, preprocessing procedures, and experimental settings are described in detail to improve the transparency and reproducibility of this study.

The specific model parameters are shown in Table 5.

Table 5. Model parameters.

Parameters	Description	Numerical value
Max seq len	Maximum sequence length	256
Learning rate	Learning rate	1e-5
Weight decay	Weight decay	0.01
Epoch	Number of training epochs	100
Batch size	Training batch size	4
Optimizer	Optimizer	Adam

4.2. Model evaluation

To evaluate the proposed two-stage knowledge extraction framework, EBERT-BiLSTM-CRF was evaluated for named entity recognition. For entity-relation extraction, the complete extraction configurations combining upstream entity recognition with downstream relation identification were evaluated against task-specific baseline configurations.

Five 8:1:1 training-validation-test splits were independently generated and reused across all proposed and baseline models. Within each run index, all models used identical training, validation, and test samples, while model training and parameter initialization were conducted independently. Results obtained from the same data split were treated as matched observations in the paired statistical analyses. Mean evaluation results across the five runs are reported.

All experiments were conducted in Python with PyTorch. Identical computational conditions and parameter configurations were applied to the training and testing of every model, allowing their performance to be compared consistently. Entity recognition and entity-relation extraction were evaluated using precision, recall, and F1-score, whose definitions are provided below [23].

$$P_{\text{recision}} = T_p / (T_p + F_p) \quad (20)$$

$$R_{\text{ecall}} = \frac{T_p}{T_p + F_N} \quad (21)$$

$$F_{1\text{-score}} = 2 \frac{P_{\text{recision}} R_{\text{ecall}}}{P_{\text{recision}} + R_{\text{ecall}}} \quad (22)$$

where T_p denotes correctly identified positive instances, T_N denotes correctly identified negative instances, F_p denotes negative instances incorrectly classified as positive, and F_N denotes positive instances incorrectly classified as negative.

4.3. Comparison of training errors

In this study, the training-error trend of the model over 40 training iterations is recorded, as shown in Figure 7. To improve model convergence efficiency, Adam is selected as the optimization algorithm for both the EBERT-BiLSTM-CRF model proposed in this study and the benchmark model used for comparison.

Figure 7 shows that the proposed model reaches a lower final loss and converges slightly faster

than the baseline. This pattern indicates more efficient feature learning for entity recognition. The faster reduction in training loss and improved convergence support subsequent knowledge extraction and fault-disposition processing in distribution networks.

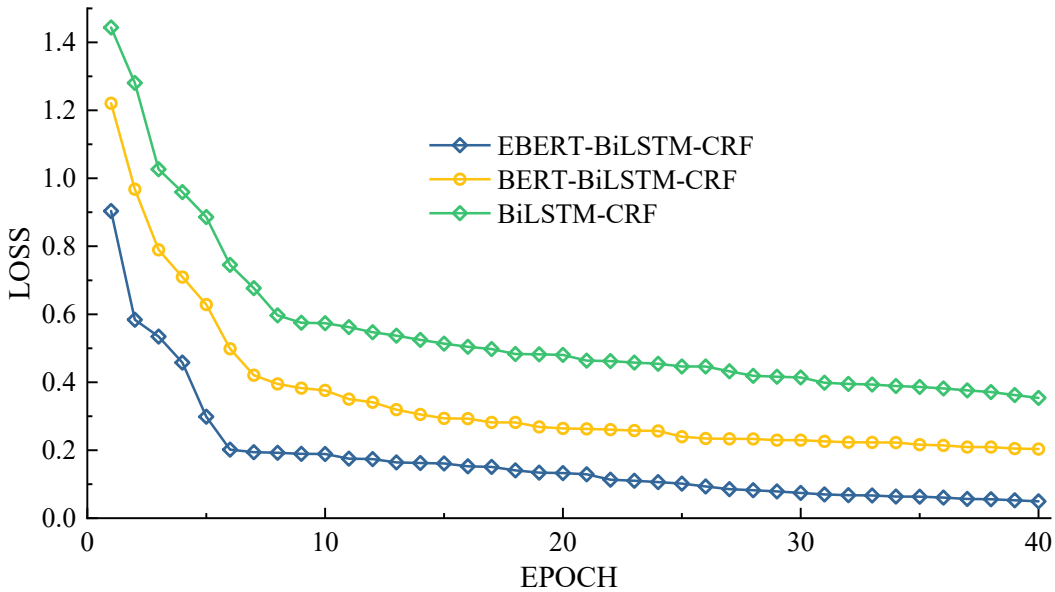


Figure 7. Training loss curves of different entity recognition models.

4.4. Comparison of different categories of named entity recognition

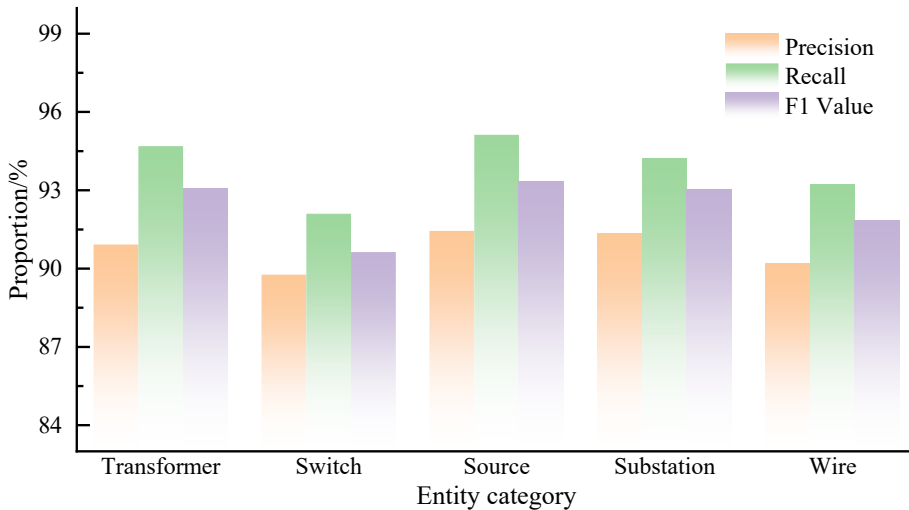


Figure 8. Entity recognition results for different entity categories.

Figure 8 shows that the EBERT-BiLSTM-CRF model developed in this study achieves a precision of 91.44% and an F1-score of 93.33% for power-supply-related named entity recognition, representing the best performance among all entity categories. By contrast, the switch category has the lowest precision and F1-score, at 89.76% and 90.63%, respectively. The performance differences among

entity categories are mainly attributable to the relatively small number of power-supply-related entities in DN fault disposition texts, whose descriptions are generally standardized and simple. In contrast, switches, cables, lines, and other entity categories occur more frequently in the text and are often embedded in complex statements, resulting in greater variation in recognition performance.

4.5. Experiment and result analysis of entity relationship extraction

The proposed and baseline models were assessed on the same five matched data splits. Table 6 reports the mean precision, recall, and F1-score from the five runs. All models followed the same entity-relation extraction protocol. A prediction was considered correct only when both the entity pair and its associated relation type were identified correctly. The individual F1-scores from the five matched runs are listed in Table 7.

Table 6 shows that the proposed BiLSTM-CRF-GA configuration achieves the highest overall performance, with a mean precision of 91.16%, a mean recall of 92.94%, and a mean F1-score of 92.06%. Compared with BiLSTM-CRF, the proposed configuration improves precision, recall, and F1-score by 7.04, 13.89, and 11.06 percentage points, respectively. Compared with LSTM-CRF, the corresponding improvements are 11.88, 16.52, and 14.58 percentage points, respectively. Relative to GlobalPointer, the proposed configuration achieves improvements of 3.78, 3.80, and 3.81 percentage points in precision, recall, and F1-score, respectively. Compared with TPLinker, the corresponding improvements are 5.74, 5.76, and 5.77 percentage points, respectively.

These results indicate that, within the present dataset, the gated attention mechanism helps suppress irrelevant contextual information and enhance the extraction of relation-specific semantic features. The mechanism represents an application-specific adaptation of gated attention rather than an entirely new attention architecture. Its contribution lies in adapting feature selection to entity-relation extraction from DN fault disposition texts. The reported results were averaged over five matched data splits.

In the experimental comparison, BiLSTM-CRF-GA denotes the complete entity-relation extraction configuration. The BiLSTM-CRF component identifies entities and determines valid entity-label sequences, whereas the gated attention component captures relation-relevant contextual information and predicts semantic relations between the identified entities. Therefore, the CRF layer is used for entity-sequence decoding rather than for relation classification.

Table 6. Performance comparison of different knowledge extraction configurations.

Model name	Precision (%)	Recall (%)	F1-score (%)
BiLSTM-CRF-GA	91.16	92.94	92.06
GlobalPointer	87.38	89.14	88.25
TPLinker	85.42	87.18	86.29
BiLSTM-CRF	84.12	79.05	81.00
LSTM-CRF	79.28	76.42	77.48

Note: The reported values are means obtained from five matched data splits and are expressed as percentages. Precision, recall, and F1-score were calculated from the unrounded experimental results and are presented to two decimal places.

To facilitate independent verification of the reported mean F1-scores, confidence intervals, and paired statistical tests, the individual F1-scores obtained from the five matched data splits are provided in Table 7.

Table 7. Individual F1-scores (%) obtained from five matched data splits.

Model name	Run 1	Run 2	Run 3	Run 4	Run 5	Mean	95% CI
BiLSTM-CRF-GA	91.45	92.55	91.90	92.50	91.90	92.06	± 0.57
EBERT-BiLSTM-CRF	87.10	87.80	87.30	87.55	86.50	87.25	± 0.61
GlobalPointer	88.56	87.98	88.68	88.07	87.96	88.25	± 0.43
TPLinker	86.60	86.02	86.72	86.11	86.00	86.29	± 0.43
BiLSTM-CRF	81.39	80.55	81.24	80.81	81.01	81.00	± 0.42
LSTM-CRF	77.82	77.20	77.94	77.32	77.13	77.48	± 0.46

Note: The same run index corresponds to the same training-validation-test split across all models. Model training and parameter initialization were conducted independently for each model. Mean values and 95% confidence intervals were calculated from the five runs using Student's t distribution with four degrees of freedom.

4.6. Statistical significance analysis

To examine performance differences among representative configurations of the two-stage framework, three configurations were selected for comparison: BiLSTM-CRF as the baseline configuration, EBERT-BiLSTM-CRF as the contextual-embedding-enhanced configuration, and BiLSTM-CRF-GA as the gated-attention-enhanced configuration. In this comparison, the model names denote alternative configurations within the complete extraction pipeline, and the reported F1-scores are relation-level results rather than the entity-recognition scores reported in Figure 8. All three configurations were evaluated using the same five training-validation-test splits and the same relation-level evaluation protocol. Results were paired according to the common data split represented by each run index. The two-sided 95% confidence intervals for the mean F1-scores were calculated using Student's t distribution with four degrees of freedom.

As shown in Figure 9, BiLSTM-CRF-GA achieves the highest mean F1-score of 92.06%, compared with 81.00% for BiLSTM-CRF and 87.25% for EBERT-BiLSTM-CRF. The corresponding improvements are 11.06 and 4.81 percentage points, respectively. The two-sided 95% confidence intervals are ± 0.42 for BiLSTM-CRF, ± 0.61 for EBERT-BiLSTM-CRF, and ± 0.57 for BiLSTM-CRF-GA. These intervals quantify the uncertainty in the estimated mean F1-scores across the five matched runs.

Paired t-tests and exact two-sided Wilcoxon signed-rank tests were used to assess the observed differences. Results with the same run index were paired because they were obtained from the same training-validation-test split. The paired t-tests yielded p-values below 0.001 for both comparisons. In contrast, the exact two-sided Wilcoxon signed-rank tests yielded a p-value of 0.0625 for each comparison and did not reach the significance threshold of 0.05. As only five matched observations were available, the difference between the parametric and nonparametric results should be interpreted cautiously. The individual F1-scores are provided in Table 7.

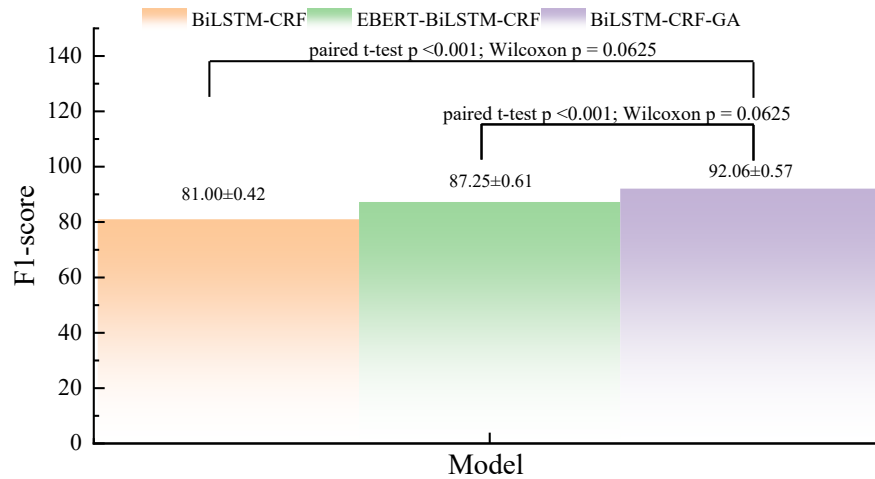


Figure 9. Statistical comparison of the mean F1-scores and 95% confidence intervals for three representative knowledge extraction configurations across five matched data splits.

5. Visualization and application of KG

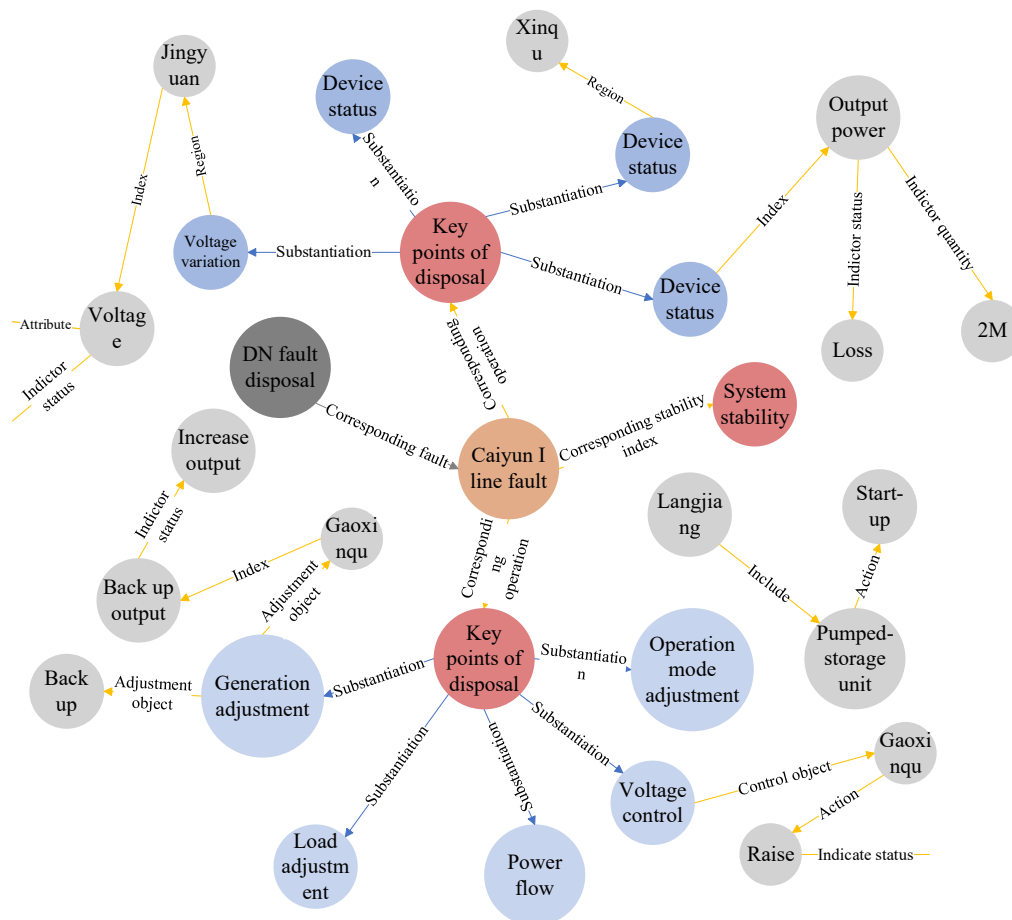


Figure 10. Partial visualization of the KG for fault handling in DN.

This study uses DN data from a region in Guangdong Province. After manual processing according to the procedure described above, the resulting triples were imported into the GridGraph graph database for storage and representation. The KG contains more than 18,700 entities and 10,400 relationships. Figure 10 presents a partial visualization of the KG.

The application of the KG to fault disposition in the DN is mainly reflected in the following two aspects:

(1) Intelligent information retrieval: Conventional keyword-based search methods for fault disposition rely on decomposition and matching and therefore lack semantic understanding. In contrast, a KG captures fault-related knowledge in a structured form and explicitly represents semantic relationships. It enables user queries to be transformed into precise concepts or entities and, through rich semantic interconnections, provides more accurate and context-aware search results.

(2) Supporting fault diagnosis: Traditional DN fault diagnosis depends heavily on dispatchers' expertise and requires real-time analysis of grid conditions and parameter variations to determine fault causes through manual reasoning. The KG for grid fault management records key features of different foreseeable faults. When a fault occurs, the system infers possible causes by referring to changes in operating states and retrieving relevant entries from the KG, thereby providing knowledge-based assistance and reducing reliance on dispatcher expertise. Furthermore, the results of each fault diagnosis are fed back as newly acquired knowledge to iteratively refine and expand the KG, enabling increasingly accurate, complete, and adaptive decision-making over time.

Figure 11 summarizes the KG-assisted fault-disposition procedure. Once a fault is detected in the DN, the required alarm and operating-state data are retrieved from the supervisory control and data acquisition (SCADA) system and the energy management system (EMS). The alarm information is parsed to generate a report describing changes in the DN's operating mode after the fault. The GridGraph database is then queried to determine whether the KG contains a fault type consistent with the parsed post-fault operating mode. If so, the fault type and corresponding key points for fault disposition are returned; if not, DN maintenance personnel analyze the fault and query the KG for the corresponding fault type. If a match is found, the system provides the identified fault category and corresponding fault disposition information. Otherwise, DN maintenance personnel must determine the nature of the fault manually and consult the KG for appropriate fault disposition guidance.

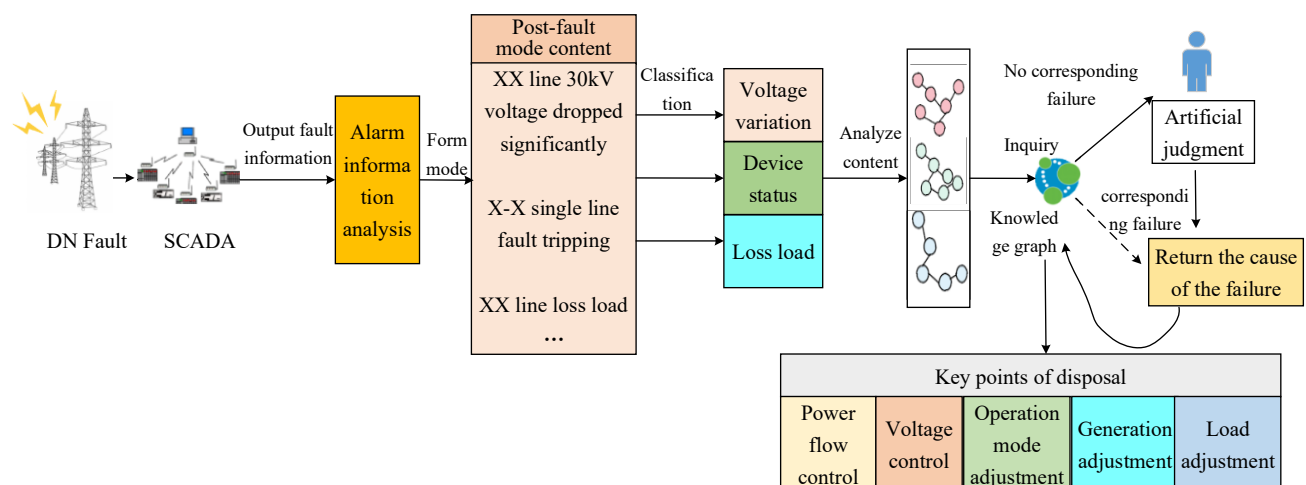


Figure 11. DN fault disposition process.

6. Conclusions

This study proposes a method for constructing and applying a DN fault disposition KG from multi-source heterogeneous data. The method addresses the data diversity and structural complexity encountered in DN fault disposition. The main conclusions are as follows:

(1) For DN fault disposition and multi-source heterogeneous data processing, this study analyses data characteristics and develops a KG construction method that combines top-down and bottom-up strategies. The implemented EBERT-BiLSTM-CRF entity recognition model converges rapidly and maintains accurate recognition across diverse DN entity categories.

(2) The BiLSTM-CRF-GA configuration is used for knowledge extraction, achieving a mean precision of 91.16%, a mean recall of 92.94%, and a mean F1-score of 92.06% across the five matched data splits. Compared with BiLSTM-CRF, it improves the mean F1-score by 11.06 percentage points, indicating improved extraction of fault-disposition knowledge from distribution-network texts.

(3) The visualization and application of the DN fault disposition KG are implemented using GridGraph as the underlying graph database, thereby providing knowledge-based auxiliary support for fault-disposition decision-making.

Although the framework performs satisfactorily in extracting DN fault-handling knowledge, its evaluation is based on practical fault data from a specific region. This limitation may restrict the assessment of the model's generalization capability. Future work will use publicly available or anonymized benchmark datasets to evaluate robustness and transferability and will consider multi-region, multi-source fault data to improve the generality of KG construction for DN fault handling.

Use of AI tools declaration

The authors declare they have not used Artificial Intelligence (AI) tools in the creation of this article.

Author contributions

Xingyu Wang: Conceptualization (equal); Data curation (equal); Formal analysis (equal); Funding acquisition (equal); Investigation (equal). Hongpeng Liu: Methodology (equal); Software (equal); Writing-original draft (equal).

Data availability statement

The raw operational data that supports the findings of this study are not publicly available and cannot be shared by the authors because of confidentiality, data-ownership, and power-system data-security restrictions. To support research transparency and methodological reproducibility, detailed information on the data sources, data composition, annotation strategy, preprocessing procedures, data-partitioning protocol, and experimental settings is provided in the manuscript.

Conflict of interest

The authors declare that they have no competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

1. Ercan SU (2024) Power line Communication: Revolutionizing data transfer over electrical distribution networks. *Eng Sci Technol, Int J* 52: 101680. <https://doi.org/10.1016/j.jestch.2024.101680>
2. Siddique MNI, Shafiullah M, Mekhilef S, et al. (2024) Fault classification and location of a PMU-equipped active distribution network using deep convolution neural network (CNN). *Electr Power Syst Res* 229: 110178. <https://doi.org/10.1016/j.epsr.2024.110178>
3. Cui J, Ran Z, Shen W, et al. (2024) Study on multi-type flexible load control method of active distribution network based on dynamic time-sharing electricity price. *Appl Energy* 357: 122479. <https://doi.org/10.1016/j.apenergy.2023.122479>
4. Zhou Y, Lin Z, Lin J, et al. (2024) Knowledge graph learning algorithm based on deep convolutional networks. *Intell Syst Appl* 22: 200386. <https://doi.org/10.1016/j.iswa.2024.200386>
5. Nawaz R, Qamar A, Wadood A, et al. (2026) Consensus-based feature selection and Bayesian-optimized stacking ensembles for comprehensive fault diagnosis in series-compensated ultra-high-voltage power systems. *Sci Rep* 16: 18678. <https://doi.org/10.1038/s41598-026-49195-7>.
6. Albalawi H, Wadood A, Bukhari SBA, et al. (2025) Design of a novel adaptive weighted multiview deep neural network for enhanced islanding detection in smart distribution systems. *IEEE Access* 13: 90333–90344. <https://doi.org/10.1109/ACCESS.2025.3571907>.
7. Chen C, Wang T, Zheng Y, et al. (2023) Reinforcement learning-based distant supervision relation extraction for fault diagnosis knowledge graph construction under industry 4.0. *Adv Eng Inform* 55: 101900. <https://doi.org/10.1016/j.aei.2023.101900>
8. Fu Y, Zhao H, Zhang H, et al. (2026) Fault detection and diagnosis of the AHU via combining a new hybrid monitoring strategy and an improved graph convolutional network. *Energy Build* 358: 117177. <https://doi.org/10.1016/j.enbuild.2026.117177>.
9. Zhang H, Zhao H, Fu Y, et al. (2026) The class labels and spatial information based fault diagnosis of air handling unit via combining kernel Fisher discriminant analysis with an improved graph convolutional neural network. *Measurement* 257: 118622. <https://doi.org/10.1016/j.measurement.2025.118622>.
10. Li Q, Wang Y, Dong J, et al. (2024) Multi-node knowledge graph assisted distributed fault detection for large-scale industrial processes based on graph attention network and bidirectional LSTMs. *Neural Netw* 173: 106210. <https://doi.org/10.1016/j.neunet.2024.106210>
11. Mao Z, Wang H, Jiang B, et al. (2024) Graph convolutional neural network for intelligent fault diagnosis of machines via knowledge graph. *IEEE Trans Ind Inform* 20: 7862–7870. <https://doi.org/10.1109/TII.2024.3367010>
12. Xiao X, Li C, Huang J, et al. (2022) Fault diagnosis of rolling bearing based on knowledge graph with data accumulation strategy. *IEEE Sens J* 22: 18831–18840. <https://doi.org/10.1109/JSEN.2022.3201839>
13. Gao F, Li G, Gu J, et al. (2023) GridOnto: knowledge representation and extraction for fault events in power grid. *IEEE Access* 11: 58863–58878. <https://doi.org/10.1109/ACCESS.2023.3284839>
14. Ji Z, Wang X, Zhang J, et al. (2023) Construction and application of knowledge graph for grid dispatch fault handling based on pre-trained model. *Global Energy Interconnection* 6: 493–504. <https://doi.org/10.1016/j.gloi.2023.08.009>

15. Li Z, Huang Z, Guo L, et al. (2023) Cognitive knowledge graph generation for grid fault handling based on attention mechanism combined with multi-modal factor fusion. *Comput Electr Eng* 111: 108855. <https://doi.org/10.1016/j.compeleceng.2023.108855>
16. Xiao N, Peng B, Li X, et al. (2023) Research on the construction and implementation of power grid fault handling knowledge graphs. *Energy Rep* 9: 182–189. <https://doi.org/10.1016/j.egy.2023.02.073>
17. Zhao Y, Zhang B, Gao D (2022) Construction of petrochemical knowledge graph based on deep learning. *J Loss Prev Process Ind* 76: 104736. <https://doi.org/10.1016/j.jlp.2022.104736>
18. Wang Y, Cheng Y, Qi Q, et al. (2024) IDS-KG: An industrial dataspace-based knowledge graph construction approach for smart maintenance. *J Ind Inf Integr* 38: 100566. <https://doi.org/10.1016/j.jii.2024.100566>
19. Guo L, Li X, Yan F, et al. (2024) A method for constructing a machining knowledge graph using an improved transformer. *Expert Syst Appl* 237: 121448. <https://doi.org/10.1016/j.eswa.2023.121448>
20. Fei Y, Yang H, Zhao G (2022) Research on intelligent construction technology of information-driven power grid security situation knowledge graph. *IFAC-PapersOnLine* 55: 102–107. <https://doi.org/10.1016/j.ifacol.2022.05.018>
21. Tang X, Feng Z, Xiao Y, et al. (2023) Construction and application of an ontology-based domain-specific knowledge graph for petroleum exploration and development. *Geosci Front* 14: 101426. <https://doi.org/10.1016/j.gsf.2022.101426>
22. Bano S, Khalid S, Tairan NM, et al. (2023) Summarization of scholarly articles using BERT and BiGRU: Deep learning-based extractive approach. *J King Saud University-Comput Inf Sci* 35: 101739. <https://doi.org/10.1016/j.jksuci.2023.101739>
23. Zhu Y (2024) A knowledge graph and BiLSTM-CRF-enabled intelligent adaptive learning model and its potential application. *Alex Eng J* 91: 305–320. <https://doi.org/10.1016/j.aej.2024.02.011>



AIMS Press

© 2026 the Author(s), licensee AIMS Press. This is an open access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0>)