Journal of Network Intelligence ©2024 ISSN 2414-8105
Taiwan Ubiquitous Information Volume **9**, Number **3**, August **2024**

Fault Entity Identification for Power Communication Equipment Incorporating BERT and BiGRU

Zhong-Yuan Nian*

State Grid Inner Mongolia East Electric Power Co
Inner Mongolia Autonomous Region, China
nianzhongyuan0452@163.com

Yan-Chang Liu

S Jining Ganghang Taiping Port Co., Ltd, China
lyanchang2022@163.com

Xiao-Da Sun

State Grid Inner Mongolia East Electric Power Co
Inner Mongolia Autonomous Region, China
suenxiaoda@126.com

Chun-Fang Mu

State Grid Inner Mongolia East Electric Power Co
Inner Mongolia Autonomous Region, China
mucf88@163.com

Ming-Shi Cui

State Grid Inner Mongolia East Electric Power Co
Inner Mongolia Autonomous Region, China
1844848479@qq.com

Chong Wang

State Grid Inner Mongolia East Electric Power Co
Inner Mongolia Autonomous Region, China
1315835959@qq.com

*Corresponding author: Zhong-Yuan Nian
Received September 18, 2023, revised December 22, 2023, accepted Marrch 8, 2024.

Abstract. *Since a large amount of fault text information is accumulated during the maintenance and overhaul of power communication equipment systems, effective automatic identification and extraction of key entity information can help over haulers discover fault key information in a timely manner, thus improving fault handling efficiency. In this paper, we take the fault text of power communication equipment as the research object and propose a fault entity identification method of Bidirectional Encoder Representation from Transformers (BERT) and two-way gated recurrent network for power communication equipment. Firstly, a corpus construction method for the entity of power communication equipment fault is proposed to standardize the processing and labeling process for the original fault text; secondly, BERT is proposed as a word embedding model to transform the fault text into word vector, fusing bidirectional gated recurrent network (BiGRU) to semantically encode the vector sequence and Conditional Random Field (CRF) to label constraints, so as to obtain the global optimal sequence; finally, the globally optimal sequence is obtained, a comprehensive comparison is made between this model and other models by constructing corresponding evaluation indexes. The experimental results show that the proposed method is significantly better than other entity recognition methods and can identify the key entity information in the text more accurately.*
Keywords: Fault text for power communication equipment, Entity identification, Corpus construction, BERT, BiGRU, CRF

1. Introduction.

In recent years, with the deepening of smart grid construction, the rapid development of power communication system has become an important engine to promote the modernization of the energy industry. In this process, the rapid increase of power communication equipment makes fault management a challenge that cannot be ignored in smart grid operation and maintenance. A large number of fault cases accumulated in the maintenance and overhaul of power communication systems [1] provide us with valuable data resources, mostly in the form of text records the fault site points, fault alarm levels, fault causes and other information, if we can utilize advanced technology to effectively identify and extract the key information in the text, it will be of great help to the fault location and disposal of work. However, it is not easy to extract useful information from these cases and utilize it.

Past studies have highlighted the heterogeneity in the quality of fault texts for power communication equipment and the difficulty of effectively utilizing unstructured texts [2,3,4,5]. These challenges make traditional entity recognition models unable to fully satisfy our need for critical entity information. Therefore, we urgently need to conduct more in-depth research to deeply analyze the faulty texts of power communication equipment through advanced technological means, with a view to mining more detailed and useful information from them.

The main objective of this study is to utilize entity recognition techniques to accurately and efficiently extract key information from the fault text of power communication equipment, including but not limited to fault site points, fault alarm levels, and fault generation causes. With the help of advanced natural language processing and machine learning methods, we will aim to improve the efficiency and accuracy of fault management in power communication systems. By deeply analyzing the fault texts of power communication equipment, we expect to provide smarter and faster support for the maintenance and overhaul of power communication systems.

1.1. Motivation and contribution.

In order to solve the problems of difficulty in effectively processing fault texts in power communication equipment and the inability of traditional models to fully integrate the contextual information of fault texts for entity recognition, this paper proposes a power communication equipment fault entity recognition method based on BERT and bidirectional gated recurrent network, which solves the problems of fuzzy word boundaries and semantic diversity in power fault data samples. By using recognition accuracy (P), recall (R), and F1 value (F) as evaluation indicators, a comprehensive comparative analysis was conducted with other recognition models, proving that the method proposed in this paper is more effective in identifying faulty entities in power communication equipment.

1.2. **Layout of the paper.** The following is the rest of the paper's layout: The second section introduces the construction of the power communication equipment fault entity corpus. The BERT fault text word vectorization module is mentioned in the third section, the fourth section describes the BiGRU fault text semantic encoding process, and the fifth section is the CRF global optimal sequence acquisition module. The experiment and analysis will be presented in the sixth section. Finally, we'll go over the conclusion.

Figure 1 shows the general framework diagram of the model proposed in this paper, which mainly consists of three main parts. Firstly, we build a corpus of power communication equipment fault entities, and carry out text normalization, word separation, deactivation and text annotation on the original fault text data to standardize the fault text processing and entity annotation process; secondly, we pass the obtained fault corpus data into BERT to obtain fault text vector sequences, pass the obtained vector sequences into BiGRU for semantic encoding, and pass the output of BiGRU into CRF for label constraint, so as to obtain the global optimal sequence and complete the fault entity corpus of power communication equipment. The output of BiGRU is passed into CRF for label constraint to obtain the global optimal sequence and complete the construction of the fault entity recognition model for power communication equipment; finally, the entity recognition model is obtained and the corresponding evaluation index is constructed to complete the evaluation of the model performance.

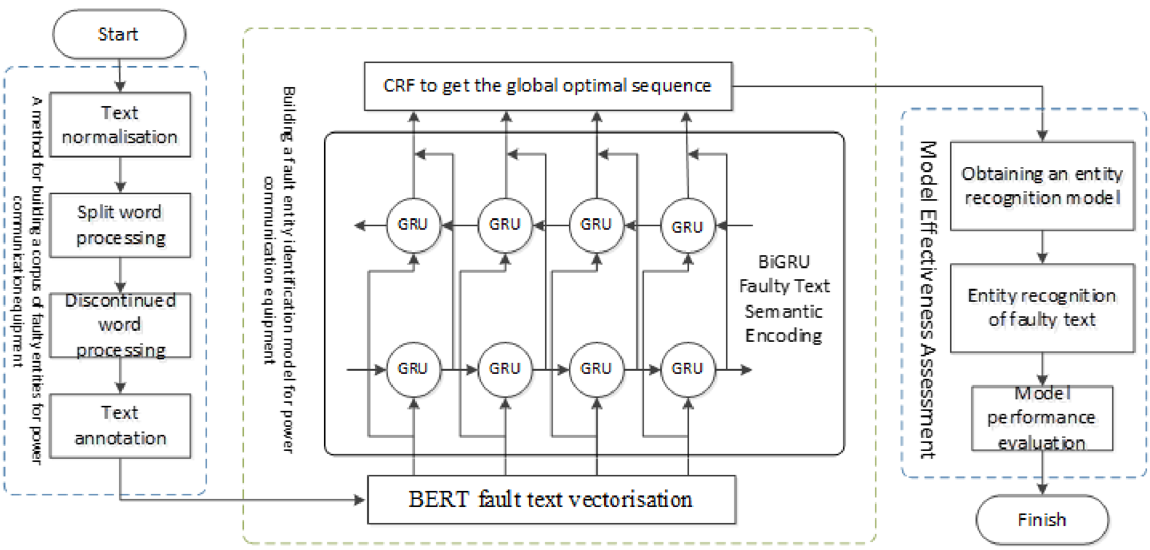

Figure 1. The overall framework of the fault entity identification method for power communication equipment

2. **Related work.** The following is a summary of the literature review for the suggested procedure as presented in this portion of the paper. Due to the rise of intelligent algorithms such as deep learning and important advances in natural language processing

techniques [5], scholars at home and abroad have made some progress in extracting textual entity information in the electric power domain [6]. The literature [7, 8] provides an overview of the concepts and methods of text entity extraction in the electric power domain and summarizes the importance of entity extraction. The literature [9, 10] applied machine learning algorithms to different areas of the electricity system respectively, but simple text classification efforts were unable to fully exploit the value embedded in electric power texts. Literature [11] constructed a dependency syntax tree for power equipment defect text and defect classification criteria text based on the dependency syntax analysis technique, which accomplished the accurate identification of defect information from the text, but the migration capability of this model is weak. The literature [12] achieved the mining of transformer operation and maintenance text information by constructing a deep semantic learning framework, but the learning performance of the model was poor. The literature [13, 14, 15] proposed a convolutional neural network-based model for mining power-related textual information and performing whole-life state evaluation of circuit breakers, but measures to improve the generalization capability of the network model were not considered. Literature [16] provides formal and informal analysis by improving the PAuth scheme thus enabling secure transmission of information between different entities. Literature [17] proposes a two-phase training strategy for grouped sparse connections to improve the efficiency of model training. Literature [18] combines QGA and LVQ neural networks and uses quantum bit coding for text update evolution. Literature [19] uses an improved VDERSc scheme to achieve forward privacy protection of data. Literature [20] Providing more reliable information through blockchain technology. Literature [21] adopts ASP as access ASP is used as the access structure to effectively realize revocable and fine-grained information access control. The literature [22] uses G-ABEET scheme to realize one-to-many data sharing. The literature [23] uses the G-ABEET scheme to realize one-to-many data sharing, and the literature proposes the FABRIC scheme, which are both helpful for the access of power fault information. The literature [24] used a bidirectional long- and short-term memory network for grid text mining, but the model could not fully incorporate contextual information for text information mining. The literature [25] used BiLSTM-CRF to extract entity information from power defect texts, but the word separation process still brought error accumulation to the model.

In summary, in order to solve the problems that the fault text of electric power communication equipment is difficult to be processed effectively and the traditional model cannot fully combine the fault text context information for entity recognition, this paper proposes a BERT and two-way gated recurrent network for the entity recognition of power communication equipment faults. Firstly, the entity corpus of electric power communication equipment faults is constructed, and the processing process of fault text and entity labeling are normalized; secondly, BERT is used as a vector embedding layer to obtain vector sequences of fault text, and the obtained vector sequences are input to BiGRU for semantic encoding, combined with CRF for labeling constraints, so as to complete the construction work of entity recognition model; finally, the Through experimental validation, the recognition accuracy (P), recall (R), and Fl value (F) are used as evaluation indexes, and a comprehensive comparative analysis is conducted with other recognition models to prove that the method proposed in this paper is more effective for the recognition of faulty entities in power communication equipment.

3. **Corpus Construction of Faulty Entities for Power Communication Equipment.**

3.1. **Power communication equipment fault text characteristics.** Compared to ordinary Chinese text messages, faulty text contains the following main features.

(1) Some of the words in the fault text have a high degree of similarity, such as "half-duplex", "full-duplex", so they need to be analyzed in conjunction with the semantic information in the context.

(2) The fault text contains a large number of specialized terms, such as "optical module", "laser", "clock board", etc. Traditional natural language techniques, such as lexical annotation, text separation, etc. cannot be directly applied to the fault text.

(3) The structure of the entity words of the fault site in the fault text is also relatively diverse such as "STM16 optical physical interface", "Ethernet OTH system port", etc. The entity words are mixed with alphabetic and numeric information, which is more difficult to identify.

3.2. **Corpus construction process.** With full consideration of the fault text characteristics of power communication equipment, the entities in the corpus are mainly divided into four major categories, namely fault information entities (FAULT), fault site point entities at the hardware level (PART), fault equipment entities (DEV), and fault site point entities at the network level (TERMINOLOGY), with the following main operational processes:

(1) Text content normalization. Through data pre-processing to standardize the fault text, such as the alarm severity level there is "Minor". "minor" case format is not uniform. Therefore, data pre-processing is used to improve the quality of fault text, reduce the noise data of fault text, and improve the accuracy of NER.

(2) Word separation processing. Since Chinese words are not separated by spaces, it is necessary to separate words from the text. In this paper, we build a fault dictionary for electric communication equipment and use HMM model and the Viterbi algorithm to separate words from the text.

(3) Removal of deactivated words. In order to improve the accuracy of text annotation and reduce the noise data in the text, this paper carries out the deactivation word removal work. In this paper, a special deactivation dictionary for power communication equipment faults is constructed by adding the corpus data of power communication equipment faults on top of the general Chinese deactivation word list.

(4) Faulty text annotation. In this paper, the corpus annotation process adopts the "BIO" annotation mechanism, where B represents the first word of the entity, I represents the middle and last word of the entity, and O represents the non-entity elements. The label definitions of the entity objects in the fault text are shown in Table 1. The labeling samples are shown in Table 2.

Table 1. Definition of labels for entity objects

Marking style	**Meaning**
B-FAULT	Fault message initials
I-FAULT	Fault message first remaining word
B-PART	Fault area braille initials
B-DEV	Power communication equipment initials
I-DEV	Remaining words for power communication equipment
B-TERMINOLOGY	First words of technical terms
I-TERMINOLOGY	Terminology remaining words
O	Other entities

Table 2. Sample markers

Text	Marking	Text	Marking
交	B-DEV	物	B-FAULT
叉	I-DEV	理	I-FAULT
板	I-DEV	故	I-FAULT
出	O	障	I-FAULT
现	O	。	O

4. **BERT fault text word vectorization module.** BERT is an advanced pre-trained word vector model proposed by Devlin et al. [26], a Google team, which is a Transformer-based deep bidirectional language representation model, characterized by its ability to fully integrate the fault context for pre-training. The pre-training process of the BERT model consists of two main tasks: Masked Language Model (MLM) and Next Sentence Prediction (NSP).

4.1. **BERT pre-training process.** The mask language model is used to predict the vocabulary of Mask by randomly Masking certain words of the sentence in the faulty text, generally defaulting to 15% of the vocabulary in the Mask sentence, thus allowing the Transformer to be used to predict the vocabulary of the Mask in combination with the information of the faulty context, as shown in Figure 2. For the input fault information, the algorithm randomly masks the characters in the sentence so that the masking result is predicted in conjunction with the context, thus allowing the BERT model to more fully understand the specific meaning of a character in the sentence.

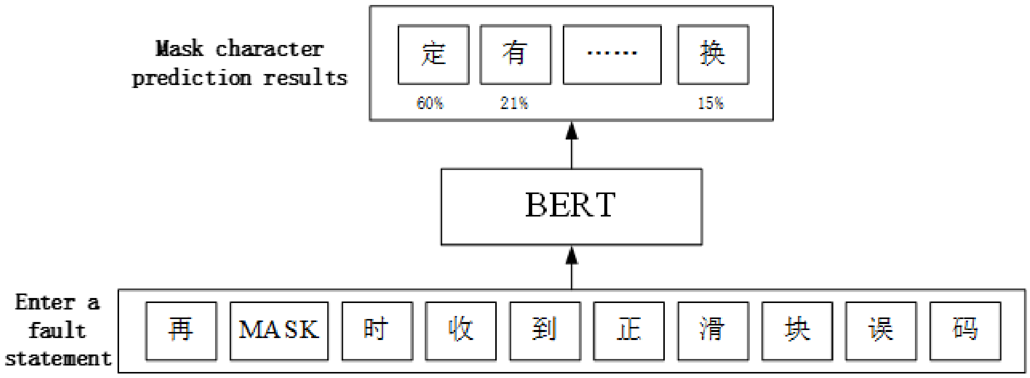

Figure 2. Mask language model

In the entity recognition task for faulty text, it is not only the contextual information between words needs to be analyzed, but also the relationship between sentences needs to be understood and reasoned, so the BERT model constructs a binary classification for the next sentence prediction task, and the specific steps are that before each training, sentence A and sentence B are randomly selected from the faulty text, 50% are the correct adjacent sentences, 50% are a randomly selected sentence, by extracting the semantic features of both sentences to make a prediction whether it is the next sentence or not. As shown in Figure 3, a "CLS" identifier is added to the first part of the sentence, and "SEP" is used to separate the sentences. There is a relationship between "temperature control system is damaged" and "BERT", so the result of BERT is 1.

4.2. **Vector coding process.** The most important component of BERT is the bi-directional Transformer coding structure, and the specific structure of the Transformer coding unit is

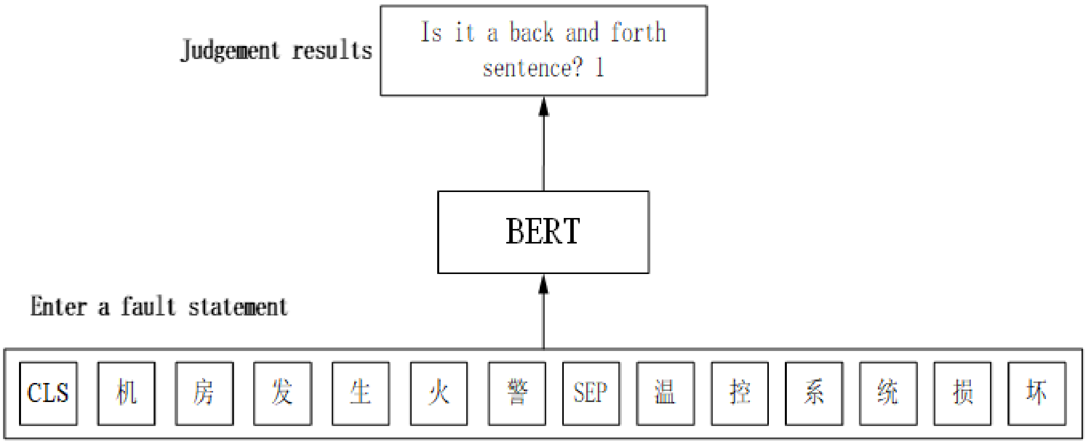

Figure 3. Next sentence prediction task

shown in Figure 4. The core module in the encoding unit is the Self-Attention mechanism, as shown in Equation (1):

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{1}$$

Among the Q, K and V are the input word vector matrices and d_k is the input vector dimension. The interrelationships of each word in the sentence for all words in the sentence are first calculated, and then using these interrelationships to adjust the importance (weights) of each word one can obtain a new representation of the semantic features of each word. This new representation implies not only the word itself, but also the relationship of other words to this word [27], so that the transformer semantic encoding has more global semantic representation information than a simple word vector.

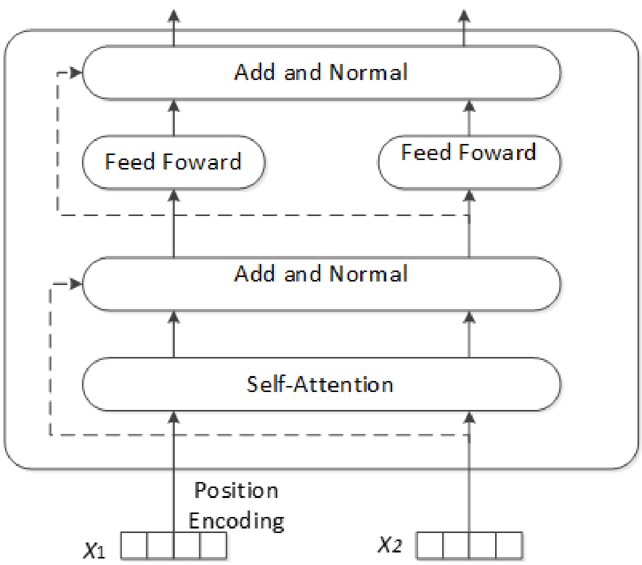

Figure 4. Transformer coding unit

In order to extend the ability of the model to focus on different locations and to increase the representation subspace of the attention unit, the Transformer uses a "multi-head" model, as shown in Equations (2) and (3):

$$\text{MultHead}(Q, K, V) = \text{Concat}(\text{head}_1, \ldots, \text{head}_h)W^O \tag{2}$$

$$\text{head}_i = \text{Attention}(QW_i^Q, KW_i^K, VW_i^V) \tag{3}$$

In the processing of the fault text, the meaning of the words in different positions in the sentence will be different, such as "the backup clock board can not lock the main clock board clock", so the Transformer module uses the location embedding method to add location features, as shown in Equation (4), Equation (5):

$$PE(pos, 2i) = \sin\left(\frac{pos}{10000^{\frac{2i}{d_{\text{model}}}}}\right) \tag{4}$$

$$PE(pos, 2i+1) = \cos\left(\frac{pos}{10000^{\frac{2i}{d_{\text{model}}}}}\right) \tag{5}$$

5. **BiGRU Faulty Text Semantic Encoding.** Since in the process of entity recognition for faulty text, there is a certain semantic correlation between words in faulty text, this paper introduces a Gate Recurrent Unit (GRU) model with a gating mechanism to perform deep feature extraction work on the input faulty text vector. The gate-controlled recurrent unit is simplified from the long-short memory neural network, i.e., the forgetting gate and the input gate are combined into one update gate, which has a simpler structure, fewer parameters, and more efficient training compared with the Long Short-Term Memory (LSTM)]28, 29], and can complete the semantic coding work for faulty text more effectively. The structure of the GRU model is shown in Figure 5, and x_t which represents the input at the moment, the z_t represents the update gate, and r_t is a reset gate for controlling information loss, and h_t is the implicit state, where x_t contains the BERT word embedding and the corresponding word feature embedding.

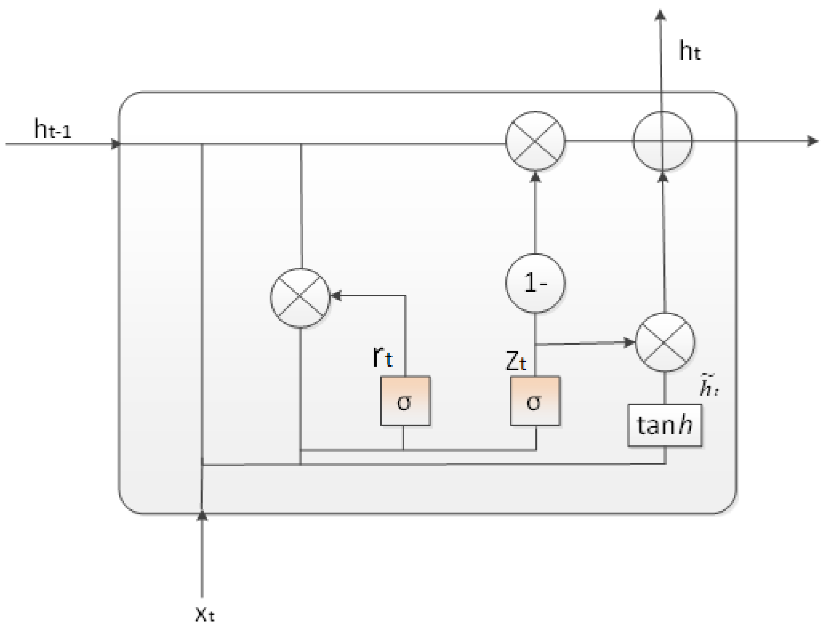

Figure 5. GRU coding unit

The update gate is used to control how much of the previous state information is brought into the current state. The larger the value of the update gate, the more the previous state information is brought in. The reset gate is used to control the degree of ignoring the state information of the previous moment, and the smaller the value of the reset gate, the more the ignoring. In the calculation of GRU, Equation (6)-Equation (9) is usually used to calculate the information controlled by each gate.

$$z_t = \sigma\left(W_i * [h_{t-1}, x_t]\right) \tag{6}$$

$$r_t = \sigma\left(W_r * [h_{t-1}, x_t]\right) \tag{7}$$

$$h_t = \tanh\left(W_c * [r_t \cdot h_{t-1}, x_t]\right) \tag{8}$$

$$h_t = (1 - z_t) \cdot c_{t-1} + z_t \cdot h_t \tag{9}$$

Where x_t refers to the input at the current moment, the σ is usually a sigmoid function that adds or multiplies vectors, $\cdot$ is the dot product of the two vectors corresponding to W_i, W_r, and W_c represents the weight matrix, and $*$ denotes the product of matrices.

The standard GRU receives input as a sequence of faulty text, which can only process the forward information and ignore the backward information. BiGRU contains a GRU network of forward and backward for each input faulty sequence, as shown in Figure 6. The output of the BiGRU network is obtained by the joint action of these two GRU networks, which can be used for the input faulty word vector x_t. Extracting bidirectional semantic information $\overrightarrow{h_t}$ and $\overleftarrow{h_t}$, and finally splice the vector operation results in both directions to obtain the final operation results h_t, which ensures that the model can capture feature information from two different directions and improve the recognition effect of the model.

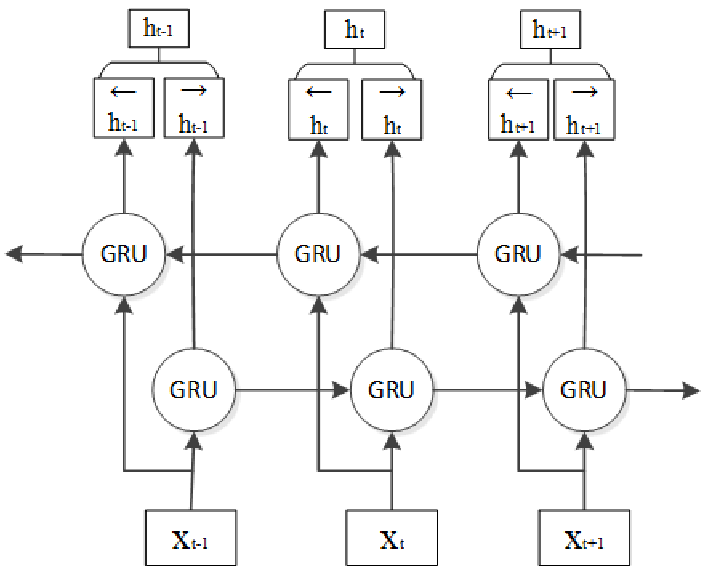

Figure 6. BiGRU fault text semantic encoding

6. **CRF global optimal sequence acquisition module.** Conditional random fields can be used for named entity recognition, syntactic analysis, lexical annotation, etc [29], as BiGRU cannot fully consider the dependencies between faulty entity labels. Therefore, in this paper, CRF is used to fully consider the connected relationships among faulty entities, so as to obtain the global optimal sequence and improve the accuracy of entity

recognition. Specifically, given $X = (x_1, x_2, \cdots, x_n)$ is the input sequence, the predicted output sequence result of the probability score is calculated as shown in Equation (10):

$$S(X, y) = \sum_{i=0}^{n} A_{y_i y_{i+1}} + \sum_{i=1}^{n} P_{iy_i} \tag{10}$$

Where the transfer matrix element $A_{y_i y_{i+1}}$ represents the probability of transfer between the semantic encoding of the y_i character of the input to the tag y_{i+1}. The transfer probability between P_{iy_i} represents the probability that the first i word tagged as y_i. The probability of the total score of correct labeling is quoted with the sum of all possible labeling scores, and the sequence path is normalized to produce the probability of labeling sequence y conditional on the input sequence X. The probability is shown in Equation (11).

$$P(y|X) = \frac{e^{S(X,y)}}{\sum_{\hat{y} \in Y_x} e^{S(X,\hat{y})}} \tag{11}$$

In the training process, the loss function is obtained by using the log-maximum likelihood estimation method to obtain the correct label sequence about y^* of log probability as shown in Equation (12).

$$\log(P(y^*|X)) = S(X, y^*) - \log\left(\sum_{\tilde{y} \in Y_x} e^{S(X,\tilde{y})}\right) \tag{12}$$

Finally, the highest score result is output according to Equation (13) as the final labeling result of the faulty entity of power communication equipment, thus completing the construction of the faulty entity identification model of power communication equipment.

$$y^* = \arg\max_{y \in Y_x} S(X, \tilde{y}) \tag{13}$$

7. Experiment and analysis.

7.1. Experimental data and indexes. The experimental data used in this paper are mainly from a state network ICT company, which mainly records the information of power communication equipment faults and fault disposal measures in the company, with a total of 3115 fault records, and the specific fault text structure is shown in Table 3.

Table 3. Description of fault text content

Properties	**Content Description**	**Properties**
Alarm number	Alarm number of the fault	Alarm severity level
Fault site point	The specific area where the fault occurred	Causes
Fault information	Specific information about the fault	Treatment
Properties	Content Description	Properties
Alarm number	Alarm number of the fault	Alarm severity level

In this paper, the accuracy (P), recall (R) and F1 value (F_1), which are commonly used evaluation metrics in entity recognition, are used as the evaluation criteria for model performance. They are defined as the following Equation (14) to Equation (16). Where T_p is the number of correct entities identified by the model, and F_p is the number of irrelevant entities identified by the model, and F_n is the number of entities that are not detected by the model.

$$P = \frac{T_p}{T_p + F_p} \times 100\% \tag{14}$$

$$R = \frac{T_p}{T_p + F_n} \times 100\% \tag{15}$$

$$F_1 = \frac{2PR}{P + R} \times 100\% \tag{16}$$

7.2. **Experimental environment and parameters.** The computer configurations and parameter configurations used in the implementation are shown in Table 4.

Table 4. Computer configuration and experimental environment

computer configuration	**Experimental environment**
Windows 10 operating system	seq_length=128
Intel(R) Xeon(R) CPU E5-2650 v3 @2.30GHz 2.30GHz	batch size=64
64GB RAM	learning_rate=default value
python 3.8	epoch=50
Tensorflow 2.5.0	
gensim 3.8.3	

7.3. **Experimental results and analysis.** The computer configurations and parameter configurations used in the implementation are shown in Table 5.

7.3.1. *Comparative analysis of model performance.* In order to verify the superiority of the proposed method in this paper, four sets of experiments were conducted to identify the four types of entities in the faulty text, and the specific comparison results are shown in Table 5.

From the experimental comparison results in Table 5 and Figure 7, we can obtain the following content. In terms of comprehensive evaluation metrics, for accuracy, recall and F1 values, the BERT-BiGRU-CRF model improves 12.07%, 13.89% and 12.99% for each metric compared with the BiLSTM-CRF model, respectively. The improvement of 11.48%, 7.37% and 9.62% for each metric compared with the BiGRU-CRF model indicates that the feature extraction capability is stronger using the BERT model, which can more fully fuse the contextual information of the faulty text for feature extraction. Compared with the BERT-BiLSTM-CRF model, the improvement of each index is 10.6%, 4.62% and 7.62%, respectively, which indicates that the semantic coding of faulty text using BiGRU can capture the features in the whole sentence at a deeper level.

Table 5. Comparison of experimental effects of models

Models	**P**	**R**	**F1**	**Time**
BiLSTM-CRF	74.34	75.00	74.45	38
BiGRU-CRF	74.93	81.52	77.82	33
BERT-BiLSTM-CRF	75.81	84.27	79.82	25
BERT-BiGRU-CRF	86.41	88.89	87.44	20

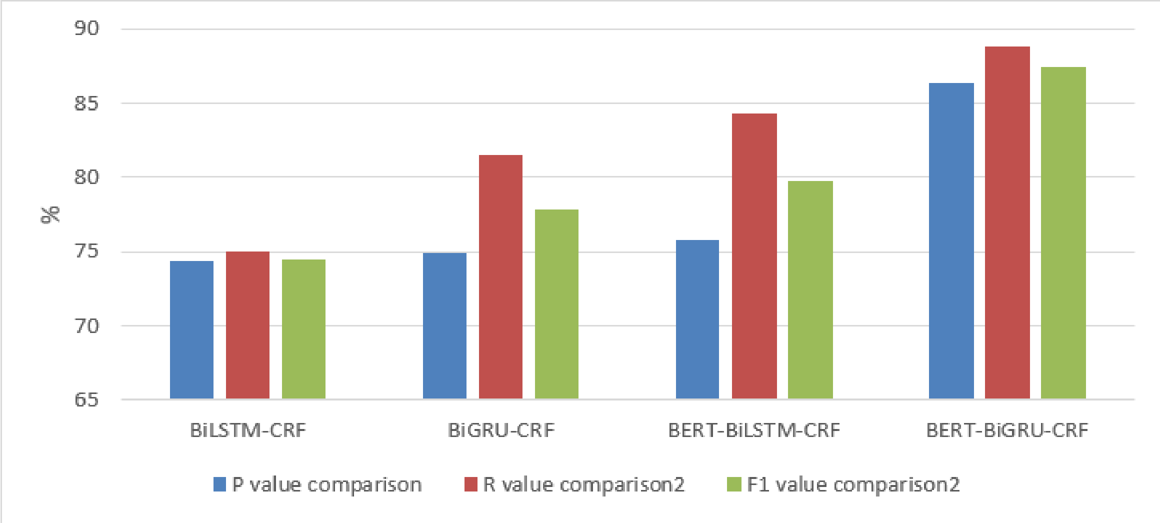

Figure 7. Comparison of various indicators of the model

From Table 6 and Figure 8 for different entity type recognition effects, comparing the Fl values of each model of BiLSTM-CRF, BiGRU-CRF, and BERT-BiLSTM-CRF can be obtained. In the recognition of fault information entities, the improvement is 9.8%, 3.66% and 5.44%, respectively; in the recognition of fault site point entities, the improvement is 7.14%, 13.56% and -0.88%, respectively; in the recognition of power communication equipment entities, the improvement is 20%, 17.73% and 16.32%, respectively; in the recognition of specialized terminology entities, the improvement is 16.69%, 13.38% and 11.61% respectively. By comparing the recognition effects of different entities, it can be seen that the advantages of this method are greater in the recognition of power communication equipment entities and professional terminology entities, but the recognition effect of the fault site entity is relatively poor, mainly because of the complex structure of the entity information composition, there are combinations of Chinese, English and numbers. In conclusion, the entity recognition model proposed in this paper has a good recognition effect, which fully proves the effectiveness of this paper's method.

In our experiments, we not only focus on the recognition performance of the models, but also analyze the computational overhead and communication cost of the models in detail. In the inference phase, we measured the average time required for each model to process one sample and observed the performance of the models in different hardware environments. First, we note that the BERT-BiGRU-CRF model performs more efficiently compared to the other models. Specifically, the average processing time of BiLSTM-CRF, BiGRU-CRF, and BERT-BiLSTM-CRF is 38 minutes, 33 minutes, and 25 minutes, respectively, while BERT-BiGRU-CRF is only 20 minutes. This shows the significant advantage of BERT-BiGRU-CRF in inference speed, which is suitable for application scenarios with high real-time requirements. Second, we analyze the hardware resources required by each model in the inference phase. The results show that BERT-BiGRU-CRF has lower GPU memory occupation compared to other models, further confirming its superiority in computational efficiency. Finally, we examine the scalability of the model on large-scale data. We find that the computational overhead of BiLSTM-CRF and BiGRU-CRF increases rapidly as the data size increases, while the BERT-BiGRU-CRF model still maintains a low computational overhead under large-scale data, demonstrating its superior scalability.

Table 6. Comparison of experimental effects of models

Entity name	**BiLSTM-CRF**	**BiGRU-CRF**	**BERT-BiLSTM-CRF**	**BERT-BiGRU-CRF**
FAULT	78.57	84.71	82.93	88.37
PART	76.19	69.77	84.21	83.33
DEV	70.00	72.73	73.68	90.00
TERMINOLOGY	71.11	74.42	76.19	87.80

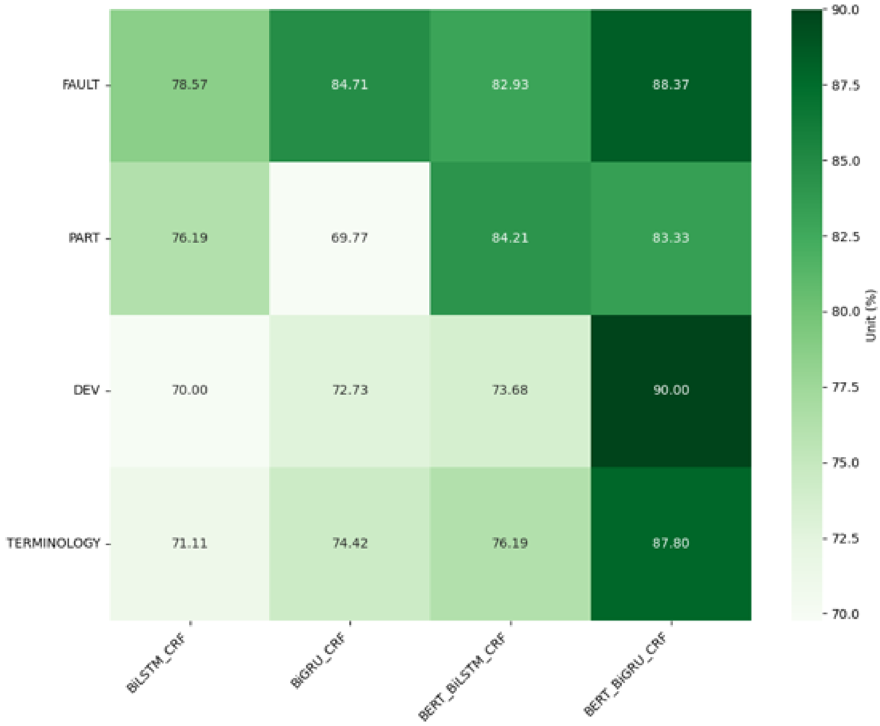

Figure 8. Comparison of F1 values for entity recognition of each model

In summary, the BERT-BiGRU-CRF model not only excels in recognition performance, but also has significant advantages in computational efficiency and communication cost, making it an ideal choice for processing large-scale data and real-time applications.

7.3.2. *Entity recognition effect demonstration.* The method proposed in this paper can effectively identify the key entity information in the fault text, and specific recognition examples are shown in Table 6. The fault text information in the table displays the entity information in the fault text (represented in bold). By comparing the contents of the recognition results, it can be seen that the proposed method can efficiently and accurately identify the various types of entity information in the fault text.

8. **Conclusion.** In conclusion, our study proposes a fusion of BERT and a BiGRU for identifying faulty entities in electric power communication equipment text. We initiate the process by constructing a dedicated entity corpus for power communication equipment faults. BERT is employed for word embedding to vectorize the fault text. The resultant word vectors are then inputted into the BiGRU layer for semantic encoding. The CRF layer, in conjunction with a state transfer matrix, outputs the optimal sequence for constructing the entity recognition model, enabling the recognition of power communication equipment faults. The identified entity information, coupled with relationship extraction techniques, is utilized to construct a knowledge map for power communication equipment faults, thereby augmenting the intelligence of fault diagnosis and maintenance.

Table 7. Example of fault text entity identification

Fault text message	Recognition results display
Physical failure of cross-board	FAULT:{Physical faults} DEV:{cross-board}
High bias currents occur in lasers in optical modules	PART:{optical modules} FAULT:{High bias current} DEV:{lasers}
Faulty external cable, faulty associated circuitry at the transceiver, mismatch of rate formats at the transceiver	PART:{the transceiver} FAULT:{Faulty external cable, faulty associated circuitry, mismatch of rate formats}
Single board circuit failure, high operating temperature of the unit	PART:{Single board} FAULT:{circuit failure, high operating temperature}

While our research has made significant strides, there are still several avenues for further exploration in future studies. Firstly, a deeper investigation into capturing complex relationships between entities could enhance the accuracy of the knowledge map. Secondly, addressing challenges related to processing multi-modal data and real-time fault detection remains an outstanding issue, warranting in-depth exploration in subsequent research. As the fields of natural language processing and neural networks continue to evolve, we will closely monitor the latest advancements to ensure our research remains at the forefront of fault diagnosis and maintenance studies.

REFERENCES

[1] Y.-Q. Song, G.-L. Zhou, and Y.-L. Zhu, "Status and challenges of big data processing technology for smart grid," Power System Technology, vol. 37, no. 04, pp. 927-935, 2013.

[2] G.-Y. Shao, H.-F. Wang, and B.-T. He, "Quality evaluation and enhancement methods of defective texts of power grid equipment," Power System Technology, vol. 43, no. 04, pp. 1472-1479, 2019.

[3] R. Guo, Q. Yang, S.-H. Liu, W. Li, X. Yuan, and X.-H. Huang, "Research and application of knowledge graph construction for power grid fault disposal," Power System Technology, vol 45, no. 06, pp. 2092-2100, 2021.

[4] H.-Y. Qu, Y.-Y. Zhang, L. Kun, S.-W. Li, X.-X. Huo, "A Knowledge-Graph-Driven Method for Intelligent Decision Making on Power Communication Equipment Faults," Electronics, vol. 12, no. 18, 2023.

[5] C. Jiang, Y. Wang, J.-H. Hu, J.-Q. Xu, M. Chen, Y.-W. Wang, and G.-M. Ma, "A deep learning-based approach to power entity information identification," Power System Technology, vol. 45, no. 06, pp. 2141-2149, 2021.

[6] Y.-X. Dai, J. Zhang, Z.-X. Ji, M.-Z. Liu, T.-L. Gao, Y.-K. Zheng, and L.-Z. Yao, "Intelligent diagnosis and auxiliary decision-making of secondary equipment in power system based on functional defect text," Electric Power Automation Equipment, vol. 41, no. 06, pp. 184-194, 2021.

[7] J.-P. Tian, H. Song, L.-F. Chen, G.-H. Sheng, and X.-C. Jiang, "A textual recognition method for equipment fault oriented knowledge graph construction," Power System Technology, vol. 46, no. 10, pp. 3913-3922, 2022.

[8] H.-Z. Tao, M.-Y. Zhai, H.-Q. Xu, X.-C. Ji, J.-B. Liu, and L.-X. Xu, "Architecture and key technologies of artificial intelligence platform adapted to application scenarios in regulation," Power System Technology, vol. 44, no. 02, pp. 412-419, 2020.

[9] J. Qiao, X.-Y. Wang, R. Min, S.-H. Bai, and T.-J. Pu, "A preliminary investigation of knowledge mapping framework and key technologies for grid dispatch fault handing," Proceedings of the CSEE, vol. 40, no. 12, pp. 178-185, 2020.

[10] G.-L. Wang, H.-S. Zhao, and Z.-Q. Mi, "Application of XGBoost algorithm in main bearing fault prediction of wind turbine," Electric Power Automation Equipment, vol. 39, no. 01, pp. 73-77+83, 2019.

[11] X.-Y. Xie, J.-H. Zhou, and Y.-J. Zhang, "Applications and challenges of deep learning in ubiquitous power IOT," Electric Power Automation Equipment, vol. 40, no. 04, pp. 77-87, 2020.
[12] G.-Y. Shao, H.-F. Wang, X.-H. Wu, J.-L. Lu, J.-H. Li, and B.-T. He, "Accurate identification method for text information of power equipment defects based on dependency syntax analysis," Automation of Electric Power Systems, vol. 44, no. 12, pp. 178-185, 2020.
[13] Y.-W. Jiang, L. Li, Z.-W. Li, C. Su, Q.-J. Wang, and X.-S. Peng, "A deep semantic learning-based textual information mining method for power transformer operation and maintenance," Proceedings of the CSEE, vol. 39, no. 14, pp. 4162-4172, 2019.
[14] Z.-Q. Liu, H.-F. Wang, J. Cao, and J. Qiu, "Research on text classification model of power equipment defects defects based on convolutional neural network," Power System Technology, vol. 42, no. 02, pp. 644-651, 2018.
[15] J. Cao, L.-S. Chen, J. Qiu, H.-F. Wang, G.-L. Ying, and B. Zhang, "Semantic framework-based text mining technology for power grid defects and its application," Power System Technology, vol. 41, no. 02, pp. 637-643, 2017.
[16] J. Qiu, H.-F. Wang, G.-L. Ying, B. Zhang, G.-P. Zou, and B.-T. He, "Text information mining technology and its application in whole-life condition evaluation of circuit breakers," Automation of Electric Power Systems, vol. 40, no. 06, pp. 107-112+118, 2016.
[17] T. Wu, Y.-Q. Lee, C.-M. Chen, et al, "An enhanced pairing-based authentication scheme for smart grid communications," Ambient Intell Human Comput, prepublish: 1-13, 2021.
[18] Y.-R. Ma, Y.-J. Peng, and T.-Y. Wu, "Transfer learning model for false positive reduction in lymph node detection via sparse coding and deep learning," Journal of Intelligent Fuzzy Systems, vol. 43, no. 02, pp. 2121-2133, 2022.
[19] F. Zhang, T.-Y. Wu, Y. Wang, R. Xiong, G. Ding, P. Mei, and L. Liu, "Application of Quantum Genetic Optimization of LVQ Neural Network in Smart City Traffic Network Prediction," IEEE Access, vol. 8, pp. 104555-104564, 2020.
[20] C.-M. Chen, Z. Tie, E.-K. Wang, et al, "Verifiable dynamic ranked search with forward privacy over encrypted cloud data," Peer-to-Peer Networking and Applications, vol. 14, no. 5, pp. 1-15, 2021.
[21] Q. Mei, H. Xiong, Y.-C. Chen and C.-M. Chen, "Blockchain-Enabled Privacy-Preserving Authentication Mechanism for Transportation CPS With Cloud-Edge Computing," IEEE Transactions on Engineering Management, doi: 10.1109/TEM.2022.3159311.
[22] H. Xiong, Z. Qu, X. Huang and K.-H. Yeh, "Revocable and Unbounded Attribute-based Encryption Scheme with Adaptive Security for Integrating Digital Twins in Internet of Things," IEEE Journal on Selected Areas in Communications, doi: 10.1109/JSAC.2023.3310076.
[23] H. Xiong, H. Wang, W. Meng and K.-H. YehSenior Member, "Attribute-based Data Sharing Scheme with Flexible Search Functionality for Cloud Assisted Autonomous Transportation System," IEEE Transactions on Industrial Informatics, doi: 10.1109/TII.2023.3242815.
[24] L. Wang, Y. Lin, T. Yao, H. Xiong and K. Liang, "FABRIC: Fast and Secure Unbounded Cross-System Encrypted Data Sharing in Cloud Computing," IEEE Transactions on Dependable and Secure Computing, doi: 10.1109/TDSC.2023.3240820.
[25] X.-M. Du, J.-F. Qin, S.-Y. Guo, and D.-F. Yan, "Text mining of typical fault cases of power equipment," High Voltage Engineering, vol. 44, no. 04, pp. 1078-1084, 2018.
[26] J. Devlin, M.-W. Chang, K. Lee, et al, "Pre-training of deep bidirectional transformers for language understanding," arXiv preprint arXiv: 1810.04805, 2018.
[27] J.-B. Li, S.-W. Fang, Y.-Q. Ren, K.-C. Li, and M.-Y. Sun, "SWVBiL-CRF: Selectable word vectors-based BiLSTM-CRF power defect text named entity recognition," 2020 IEEE International Conference on Big Data (Big Data), 2020.
[28] P. Yang, and W.-Y. Dong, "A Chinese named entity recognition method based on BERT embedding," Computer Engineering, vol. 46, no. 04, pp. 40-45+52, 2020.
[29] Q. Qin, S. Zhao, and C. Liu, "A BERT-BiGRU-CRF model for entity recognition of Chinese electronic medical records," Complexity, 2021.
[30] L.-S. Li, and Y.-K. Guo, "Biomedical named entity recognition based on CNN-BLSTM-CRF model," Journal of Chinese Information Processing, vol. 3.