electronics

Article

Named Entity Recognition for Equipment Fault Diagnosis Based on RoBERTa-wwm-ext and Deep Learning Integration

Feifei Gao [1], Lin Zhang [1], Wenfeng Wang [1], Bo Zhang [1,*], Wei Liu [1], Jingyi Zhang [1] and Le Xie [2]

[1] Air and Missile Defense College, Air Force Engineering University, Xi'an 710051, China; m18031827043@163.com (F.G.)
[2] Electronic Information School, Xijing University, Xi'an 710123, China
* Correspondence: zhb8706@163.com

Abstract: Equipment fault diagnosis NER is to extract specific entities from Chinese equipment fault diagnosis text, which is the premise of constructing an equipment fault diagnosis knowledge graph. Named entity recognition for equipment fault diagnosis can also provide important data support for equipment maintenance support. Equipment fault diagnosis text has complex semantics, fuzzy entity boundaries, and limited data size. In order to extract entities from the equipment fault diagnosis text, this paper presents an NER model for equipment fault diagnosis based on RoBERTa-wwm-ext and Deep Learning network integration. Firstly, this model uses the RoBERTa-wwm-ext to extract context-sensitive embeddings of text sequences. Secondly, the context feature information is obtained through the BiLSTM network. Thirdly, the CRF is combined to output the label sequence with a constraint relationship, improve the accuracy of sequence labeling task, and complete the entity recognition task. Finally, experiments and predictions are carried out on the constructed dataset. The results show that the model can effectively identify five types of equipment fault diagnosis entities and has higher evaluation indexes than the traditional model. Its precision, recall, and F1 value are 94.57%, 95.39%, and 94.98%, respectively. The case study proves that the model can accurately recognize the entity of the input text.

Keywords: equipment fault diagnosis; named entity recognition; RoBERTa-wwm-ext; deep learning; knowledge graph

check for updates

Citation: Gao, F.; Zhang, L.; Wang, W.; Zhang, B.; Liu, W.; Zhang, J.; Xie, L. Named Entity Recognition for Equipment Fault Diagnosis Based on RoBERTa-wwm-ext and Deep Learning Integration. *Electronics* **2024**, *13*, 3935. https://doi.org/10.3390/electronics13193935

Academic Editors: Panayiotis Bozanis and Leonidas Akritidis

Received: 12 September 2024
Revised: 23 September 2024
Accepted: 30 September 2024
Published: 5 October 2024

Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

With the extensive application of advanced science and technology in the military field, weapons and equipment continue to develop with a focus on information and intelligence. Complex high-tech equipment is constantly used in modern warfare and training, and high-level equipment's support capability is a key factor in the operational effectiveness of equipment and can determine the outcome of a war. However, the continuous application of new technology makes the complexity of equipment and the difficulty of support increase rapidly. The task of equipment support is to ensure the normal operation and use of equipment, so the main focus of equipment support is the maintenance of equipment to prevent failure. Equipment fault maintenance refers to the methods, techniques, skills, and means used for the maintenance and repair of equipment after failure. Equipment in the process of use will encounter a variety of failures, and troubleshooting and maintenance personnel mainly rely on their accumulated experience and technical specifications. In the troubleshooting process, the methods, processes, skills, and means adopted by maintenance personnel are usually recorded in the form of unstructured text, and these maintenance experiences are difficult to be effectively reused. Therefore, it is of great significance to efficiently use the data accumulated in equipment maintenance failure records, deeply mine the knowledge and potential relationship in the fault log text, and assist the maintenance personnel to quickly complete the fault analysis and troubleshooting in the process of equipment support.

Knowledge graph (KG) technology provides a better method for industry knowledge mining, representation, and management by mining the entities and relationships in the information and presenting them in the form of a visual semantic network [1]. By using the method of a knowledge graph, fault knowledge can be effectively mined from equipment fault texts. Through the integration and association of knowledge, combined with reasoning analysis, the equipment fault maintenance plan is formulated to assist maintenance personnel to quickly find, locate, diagnose, and repair faults [2]. At present, an increasing number of scholars have introduced KGs into fault diagnosis practice. For example, Deng et al. [3] explored strategies for building a KG for the fault diagnosis of robotic transmission systems. The BiLSTM network is used to capture the context information features in the fault text, the self-attention is used to accurately extract the interdependence features between characters in the multi-dimensional subspace, and the CRF is used to realize the effective identification of key entities, which plays a vital role in promoting the autonomous fault diagnosis. Tang et al. [2] used Deep Learning technology to extract the entity, relationship, and attribute information from aircraft fault diagnosis data to build a KG in the field of aircraft faults. Liu et al. [4] constructed an electrical equipment fault KG based on the text of the electrical equipment operation and maintenance records, showing the correlation of faulty equipment and components. In the process of constructing the KG of equipment fault diagnosis, NER plays a vital role. With the continuous accumulation and increase in equipment fault record data, it is unrealistic to rely on manual methods to extract the required information from a large number of texts. Deep Learning technologies have shown great potential in the task of NER for equipment fault diagnosis, which provides strong support for automatically extracting fault information and constructing KGs [5]. At present, there is little research in the field of equipment fault diagnosis NER, but researchers have carried out a lot of research and achieved certain results in the field of power equipment, aircraft, railway equipment, and the recognition of other named entities. For example, Gong et al. [6] introduced BiLSTM and CRF models on the basis of BERT, and entities related to High-Voltage Isolating circuit breakers and thermal faults are identified from power grid fault texts. Through the joint recognition model based on the sequence and TreeBiLSTM, Meng et al. [7] extracted the interdependence between the related entities and relationships in the area of aircraft health management. Using the fault reports provided by China Railway Administration as the data source, Yang et al. [8] fuses BERT, BiLSTM, and CRF to perform text mining on the fault reports of Chinese railway operation equipment and extract relevant entities.

The relevant entities in the equipment fault diagnosis text include equipment name, fault part, fault phenomenon, fault reason, and troubleshooting way. Compared with entity recognition in other fields such as finance, tourism, medicine, agriculture, culture and so on, the characters of equipment fault diagnosis entities are longer and include more proper names. In addition, there is also the phenomenon that different entity types nest with each other. At present, there is a lack of public datasets in the field of equipment fault diagnosis, and it will be a challenging work to carry out an equipment fault diagnosis NER.

In order to complete the work of NER for an equipment fault diagnosis, this study will propose an NER model based on the fusion of RoBERTa-wwm-ext and Deep Learning. Based on the equipment fault database, we extracted entities such as equipment name, fault part, fault phenomenon, fault reason, and troubleshooting way according to the fault text characteristics. Referring to NER models in other fields, this paper proposes a RoBERTa-wwm-ext-BiLSTM-CRF NER model for equipment fault diagnosis and proves the effectiveness of the model through experiments.

The main contributions of this paper are as follows:

(1) We collected and sorted out the fault diagnosis text from the equipment fault database; constructed a dedicated Chinese corpus in the field of equipment fault diagnosis; cleaned the data carefully, implemented sentence segmentation; labeled the entity labels of equipment name, fault location, phenomenon, cause, and elimination method; and completed the text preprocessing task.

(2) We use RoBERTa-wwm-ext to process the labeled fault diagnosis text, and a neural network combining BiLSTM and CRF is used to extract the context feature information of the text to obtain the optimal prediction sequence and complete the NER task.
(3) Through experiments, the performance of different models is compared, and the effectiveness of the presented NER model for equipment fault diagnosis is verified. The precision, recall, and F1 value of the model reach 94.57%, 95.39%, and 94.98%, respectively.
(4) Through experiments, the influence of different entity types on the model is evaluated, the hyperparameters of the model are explored, and the performance of NER is improved. The case study proves that the model can accurately recognize the entity of the input text.

2. Related Work

Early entity recognition is mostly based on rule templates, which are often reasonably designed by domain experts, but the expansibility of rule templates is poor and the cost of system migration is high [9]. Later, traditional Machine Learning technologies are widely applied in entity recognition. The commonly used models include ME [10], HMM [11], and CRF [12], etc. In recent years, as the core technology driving the vigorous development of AI, Deep Learning has also been widely applied in the field of natural language processing and KG and has gradually become the mainstream method of entity recognition. Compared with the early manual methods based on dictionary matching and templates, the Deep Learning model can learn features and patterns from sample data and does not need to manually select features. It has a better effect, higher efficiency, and stronger universality and is suitable for solving sequence labeling problems. RNN is suitable for processing and predicting sequence data, but it is prone to the problem of vanishing gradients when facing very long sequences. Hochreiter and Schmidhuber [13] proposed LSTM network to solve the problem of the insufficient reflux of error information and gradient attenuation. Huang et al. [14] presented a variety of sequence labeling models based on an LSTM network and proved that the BiLSTM-CRF model can effectively use forward and backward text input features through experiments. Miao et al. [15] presented a model consisting of LSTM and fully connected layers for short-term fog prediction. An et al. [16] applied a MUSA-BiLSTM-CRF model in the field of Chinese clinical NER, which greatly improved the entity recognition performance.

In 2018, Google introduced the BERT pre-training model, which quickly became a popular model in the field of NLP due to its powerful structural and semantic understanding capabilities. Since then, scholars have continuously applied pre-trained language models to named entity recognition tasks. Devlin et al. [17] utilize BERT and a bidirectional Transformer model to generate word embedding vectors containing positional information and contextual features, achieving excellent performance on sentence-level and token-level tasks. Guo et al. [18] used BERT-BiLSTM-CRF to identify case entities in Chinese domestic legal texts. Lin et al. [19] presented an entity extraction method for fault information of railway signaling equipment based on RoBERTa-wwm and Deep Learning. RoBERTa-wwm was used to generate word vectors of text sequences, and BiLSTM and CNN were used to obtain contextual features and local feature information. Liang et al. [20] proposed ALBERT fault pre-training model with fault data embedding for communication equipment faults of industrial Internet of Things. Kong et al. [21] presented a NLP algorithm based on a dictionary, language technology platform tools, and a BERT-CRF hybrid to perform entity recognition on electrical equipment fault texts in power systems and optimized the context relationship and preferred word labels. Chen et al. [22] used BERT-BiLSTM-CRF to recognize entities in the fault diagnosis text of air compressor and proved that the model showed an excellent performance in extracting entities in the field of compressor fault diagnosis by comparing with other sequence labeling models. Zhang et al. [23] used BERT-CRF to realize the recognition of power equipment fault entities and proved through experiments that the model can extract a wider range of power equipment fault entities from a small

corpus. Zhou et al. [24] used the BERT model to extract the initial semantic information in the text of power equipment defects and then further extracted the context and local semantic features through BiLSTM-CNN, which provided a reference for the intelligent extraction of power equipment text information.

At present, there are relatively few studies on named entity recognition in the field of equipment fault diagnosis. In this paper, an equipment fault diagnosis corpus is constructed, the RoBERTa-wwm-ext-BiLSTM-CRF model is applied to recognize the named entities in the equipment fault diagnosis text, and the five types of equipment fault diagnosis entities are effectively extracted: equipment name, fault part, fault phenomenon, fault reason, and troubleshooting method.

3. Methodology

In this section, we will elaborate on the NER model for equipment fault diagnosis based on the fusion of RoBERTa-wwm-ext and BiLSTM-CRF architectures, and Figure 1 shows overall structure of model, including RoBERTa-wwm-ext, BiLSTM, and CRF layer. Firstly, RoBERTa-wwm-ext layer was used to convert the equipment fault diagnosis text data into word embedding vector representation. Then, the trained word vector sequence was input into BiLSTM network layer to fully extract the context feature information in text. Finally, dependency relationship between adjacent labels was learned in CRF layer, output results of BiLSTM layer were decoded, optimal label sequence with constraints was output, the entities in sequence were extracted and classified, and NER was completed.

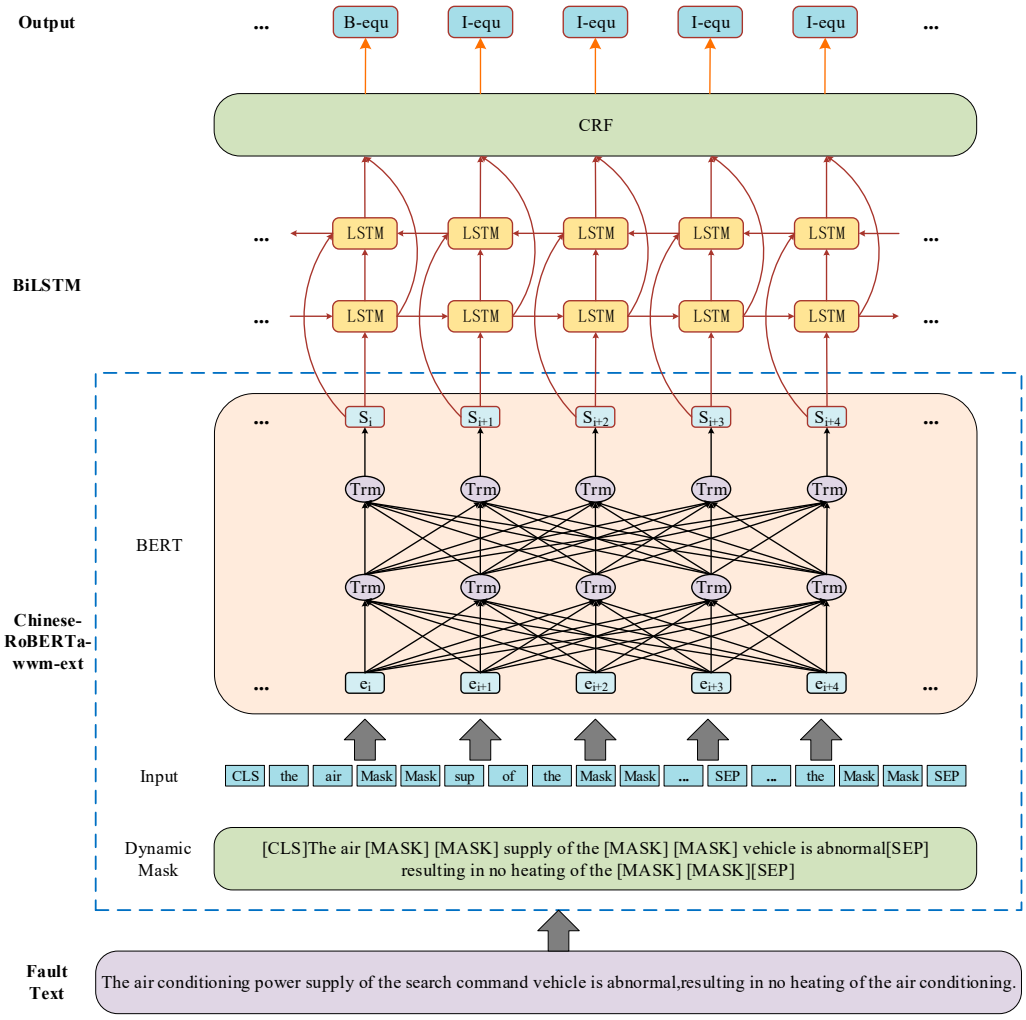


Figure 1. NER model based on RoBERTa-wwm-ext-BiLSTM-CRF network.

3.1. RoBERTa-wwm-ext Layer

RoBERTa-wwm-ext model is a Chinese pre-trained model released by HFL (HIT iFLYTEK Language Cognitive Computing Lab), which is a derivation and optimization of BERT model.

BERT model is built on the basis of the 12-layer encoder component of Transformer architecture. The deep bidirectional Transformer encoder is used to learn the rich semantic information in the text data. Figure 2 shows the framework structure of the Transformer. BERT learns the context information of the corpus through two pre-training tasks: MLM and NSP. Figure 3 shows the model structure of BERT. The input of model consists of two pieces of text concatenation, labels [CLS] for sentence classification tasks and labels [SEP] for two input sentence segmentation tasks.

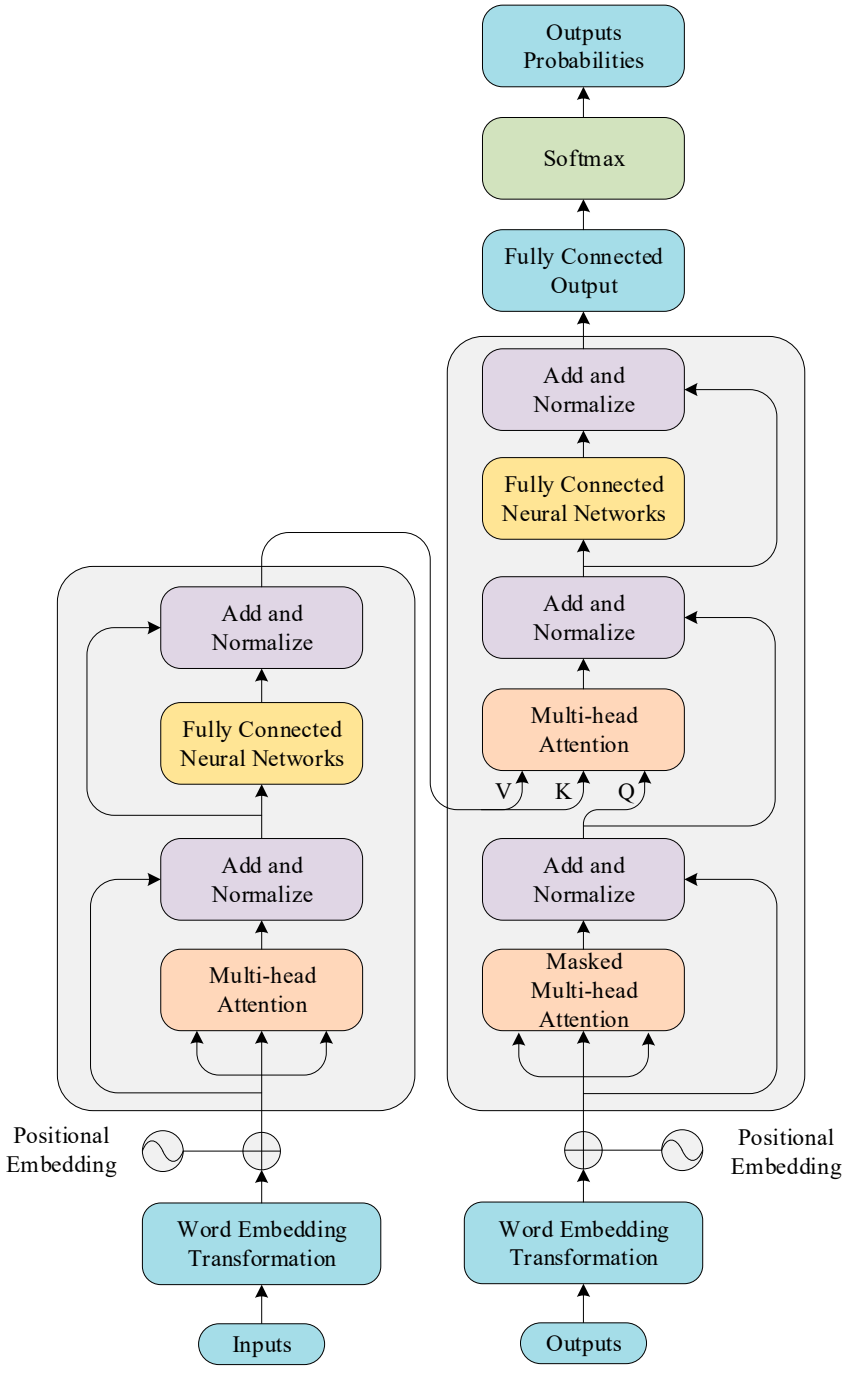


Figure 2. The structure of Transformer.

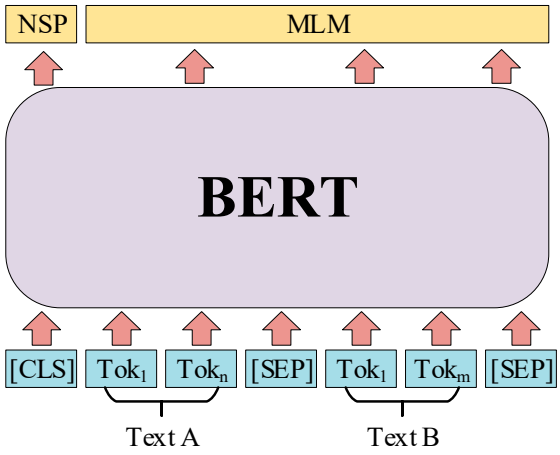


Figure 3. The structure of BERT.

Figure 4 shows the input representation of BERT, which consists of the sum of word vectors, block vectors, and position vectors [25]. The formula for computing the input representation v is as follows:

$$v = v^t + v^s + v^p \tag{1}$$

where v^t refers to the word vector, v^s refers to the block vector, and v^p refers to the position vector. The first layer of word vector represents transforming the words in the input text into 768-dimensional vectors, the second layer of block vector determines which sentence the current word belongs to, and the third layer of position vector encodes the absolute position of each word.

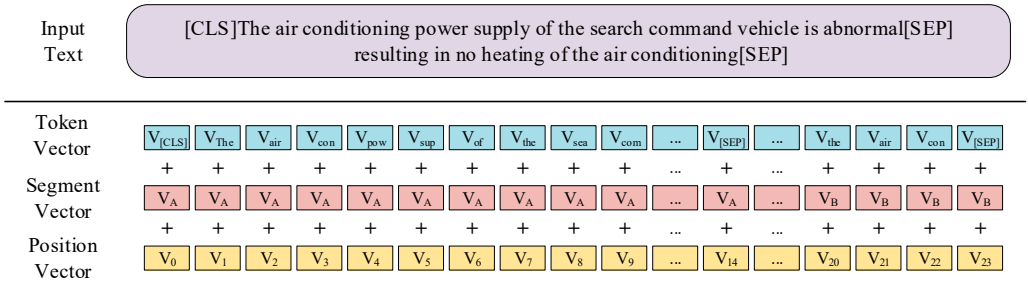


Figure 4. Input representation of BERT.

RoBERTa-wwm-ext model is an improved RoBERTa model, especially suitable for Chinese text processing. Compared with BERT model, the following improvements are made:

(1) Dynamic masking technology is introduced to ensure that the same text has different masking patterns under different training epochs, which improves the richness of training data and the efficiency of data reuse.
(2) NSP tasks are discarded to improve the efficiency of downstream tasks.
(3) Using the whole word mask technology, the context semantics can be better understood, and the accuracy and efficiency of Chinese text processing can be improved.
(4) Improving model performance by using larger batches, longer training steps, and larger data sizes.

Figure 5 shows the input representation of RoBERTa-wwm-ext model. The input text is first processed with labels [CLS] and [SEP] indicating where each text begins and ends, using a dynamic masking technique with labels [MASK] randomly masking characters in the text [19]. The input of the text consists of adding word vector, block vector, and position vector.

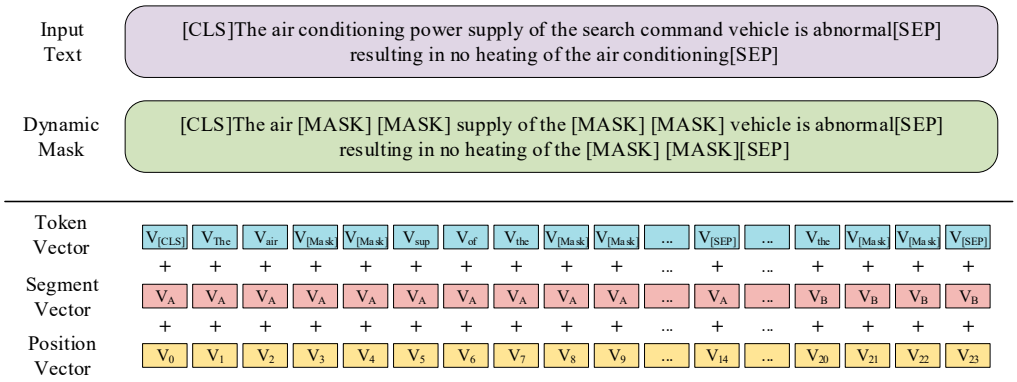


Figure 5. Input representation of RoBERTa-wwm-ext.

3.2. BiLSTM Layer

RNN is a neural network model specifically designed to process sequence data, which captures the contextual feature information of fault text through internal feedback links. However, the basic RNN has gradient explosion and gradient disappearance, which has drawbacks when dealing with long-distance dependence problems [26]. LSTM is an improvement of the basic RNN and is designed to solve the problem of long-distance dependence in sequence modeling. LSTM network consists of basic LSTM units. Figure 6 shows the internal structure of LSTM network unit.

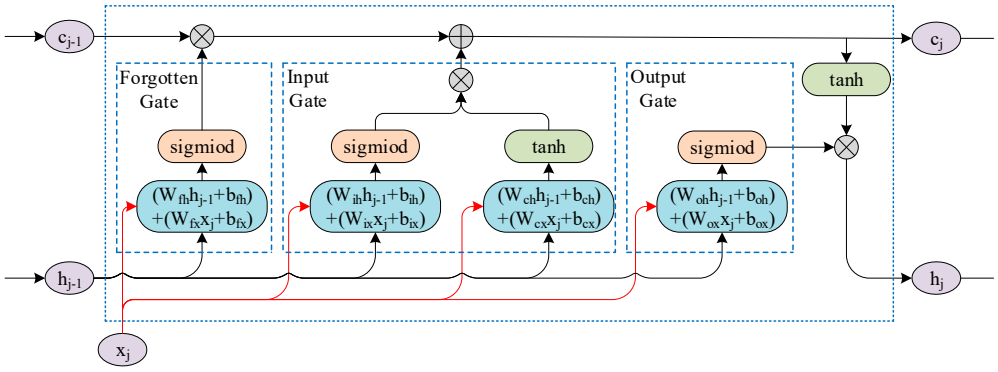


Figure 6. The internal structure of an LSTM network unit.

The mathematical expressions of LSTM network model are given by Equations (2)–(7).

$$f = sigmoid(W_{fx}x_j + b_{fx} + W_{fh}h_{j-1} + b_{fh}) \tag{2}$$

$$i = sigmoid(W_{ix}x_j + b_{ix} + W_{ih}h_{j-1} + b_{ih}) \tag{3}$$

$$g = \tanh(W_{cx}x_j + b_{cx} + W_{ch}h_{j-1} + b_{ch}) \tag{4}$$

$$o = sigmoid\hbar(W_{ox}x_j + b_{ox} + W_{oh}h_{j-1} + b_{oh}) \tag{5}$$

$$c_j = f \bigotimes c_{j-1} + i \bigotimes g \tag{6}$$

$$h_j = o \bigotimes \tanh(c_j) \tag{7}$$

The LSTM network unit processes the input data through forget gate, input gate, and output gate to realize the memory mechanism at long and short distances. Forget gate forgets the information in the memory and determines the unimportant and discarded information in the memory; Equation (2) represents its calculation process. Input gate processes the new information and decides information to be memorized; Equation (6) represents its calculation process, where Equation (3) calculates the brand new information to be memorized and Equation (4) calculates the update of the old information. Output gate

processes input information, both by direct processing of the current input and by modifying these inputs based on previously memorized information, and Equations (5) and (7) represent their computation processes, respectively.

BiLSTM is a bidirectional LSTM network constructed based on LSTM units. BiLSTM connects the same input sequence into the forward and backward LSTMS [27], concatenates the hidden layers of LSTM network, and accesses output layer together for prediction. Figure 7 shows the structure of BiLSTM network. For NER task, BiLSTM network is used to extract context feature information, which can not only realize the dependence of backward text on forward text, but also realize the dependence of forward text on backward text, which can effectively solve the dependence problem of distant entities.

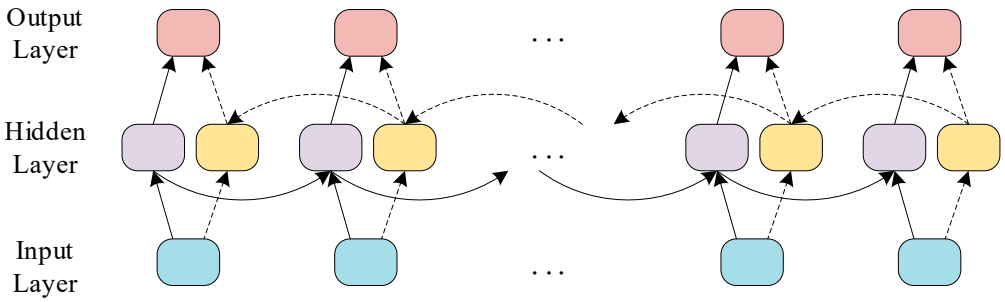


Figure 7. The structure of BiLSTM network.

3.3. CRF Layer

In NER task, BiLSTM layer extracts the context feature information of the fault text and obtains the probability of occurrence of each word on each label, but it lacks the ability to process the dependency between labels [28]. CRF can calculate the relationship between adjacent labels from a global perspective and obtain the optimal prediction sequence. CRF makes up for the BiLSTM layer's inability to deal with neighboring label dependencies and reduces the number of invalid predicted labels. The BiLSTM-CRF network model has been shown to significantly improve the precision of NER. Figure 8 shows the structure of CRF.

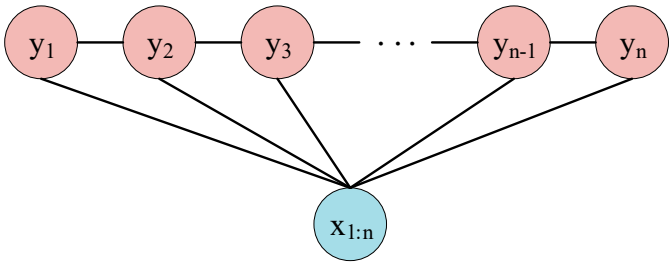


Figure 8. The structure of CRF.

In a given observation sequence $\boldsymbol{x} = (x_1, x_2, x_3, \cdots, x_n)$, a label sequence $\boldsymbol{y} = (y_1, y_2, y_3, \cdots, y_n)$ and Equation (8) calculates the corresponding $score(\boldsymbol{x}, \boldsymbol{y})$, where λ_j denotes the corresponding weight and f_j denotes the characteristic function.

$$score(\boldsymbol{x}, \boldsymbol{y}) = \sum_j \sum_i \lambda_j f_j(y_{i-1}, y_i, \boldsymbol{x}, i) \tag{8}$$

Softmax function is used to normalize all possible sequence paths, and the conditional probability distribution $p(\boldsymbol{y}|\boldsymbol{x})$ of the predicted sequence is obtained according to Equation (9), where $\widetilde{\boldsymbol{y}}$ represents the current predicted tag sequence, Y_x represents the set of all possible tag sequences, and $score\left(\boldsymbol{x}, \widetilde{\boldsymbol{y}}\right)$ represents the total score under the current

predicted sequence. Taking the logarithm of both ends of Equation (9) yields the maximum likelihood probability function of the correct predicted label, as shown in Equation (10).

$$p(\boldsymbol{y}|\boldsymbol{x}) = \frac{exp\{score(\boldsymbol{x}, \boldsymbol{y})\}}{\sum_{\widetilde{\boldsymbol{y}} \in Y_x} exp\left\{score(\boldsymbol{x}, \widetilde{\boldsymbol{y}})\right\}} \tag{9}$$

$$\log p(\boldsymbol{y}|\boldsymbol{x}) = score(\boldsymbol{x}, \boldsymbol{y}) - \log\left(\sum_{\widetilde{\boldsymbol{y}} \in Y_x} score(\boldsymbol{x}, \widetilde{\boldsymbol{y}})\right) \tag{10}$$

Finally, the maximum likelihood function $argmax()$ is used to decode, and optimal sequence is selected from all predicted label sequences, and the optimal sequence is the final label result, as shown in Equation (11).

$$y^* = argmax(score(\boldsymbol{x}, \widetilde{\boldsymbol{y}})), (\widetilde{\boldsymbol{y}} \in Y_x) \tag{11}$$

4. Experiment and Analysis

This section introduces the experimental process of the equipment fault diagnosis text NER based on the fusion of the pre-trained model and Deep Learning model, including text preprocessing, text label, experimental environment setting, training parameter setting, result analysis, and the comparison of different model combination effects.

4.1. Text Preprocessing

The original data in this paper are 1302 fault records obtained from the equipment IETM fault database and the equipment user fault register, totaling about 50,000 words. These raw texts have many problems and cannot be directly utilized. For example, most of the content exists in the form of tables and flow charts, which cannot be recognized as text. In addition, when the text length is too long, the recognition precision is often affected and decreased. Therefore, it is crucial to pre-process the raw text. In this paper, text preprocessing mainly performs text cleaning, segmentation, and format conversion. In order to obtain text for entity annotation, the main work of text preprocessing includes correcting errors, filling in missing values, cutting long text, and converting tables and flow charts into text. The standardized sentences that only include words, numbers, and punctuation are obtained by text preprocessing.

4.2. Text Label

The text label is fundamental for building corpora and performing NER tasks. In this paper, the BIO format is used to annotate the entities in the text: "B-entity type" denotes the beginning of each entity, "I-entity type" denotes the rest of each entity, and "O" denotes non-entity words. A total of 3422 entities are annotated in this paper, including five main types of equipment fault diagnosis: equipment name, fault part, fault phenomenon, fault reason, and troubleshooting way, as shown in Table 1.

Table 1. Type and quantity of equipment fault diagnosis entity.

Label	Type	Number
Equipment	equipment name	67
Part	fault part	884
Phenomenon	fault phenomenon	977
Reason	fault reason	973
Way	troubleshooting way	521

The specific labeling work has the following steps: upload the text to the Label-Studio labeling tool, set the required labels, label each entity with the corresponding label, select

the "json" format to export after the labeling is completed, and then convert the label format into the "BIO" format text through the format conversion program.

4.3. Training

In this paper, according to the allocation ratio of 80% training set, 10% test set, and 10% validation set, the dataset is divided into three parts containing 2737, 349, and 336 entities, respectively, for training, testing, and verifying model performance. The pre-trained models used in the experiments are "Chinese-RoBERTa-wwm-ext" and "Chinese-BERT-base", which contain 12 layers of transformers, 12 self-attention mechanisms, and 768 hidden layer dimensions. Word vectors output by model are the weighted average of the 12-layer network, and the final 768-dimensional word vector will be fed into the BiLSTM + CRF layer [5]. Table 2 shows experimental environment setup.

Table 2. Experimental environment.

Type	Configuration
Hardware configuration	CPU: 13th Gen Intel Core i7-13620H GPU: NVIDIA Tesla P40 OS: Windows 11 Video memory: 24 GB
Software environment	CUDA: 10.2 Python: 3.12.2 Tensorboard: 2.16.2 Transformers: 4.42.3 Tqdm: 4.66.4 Numpy: 1.26.4

To ensure reliability of results when conducting comparative experiments, it is necessary to use fixed hyperparameters for training, and the specific hyperparameters of the model settings are shown in Table 3. "max_seq_length" specifies maximum length of the input sequence, "epoch" specifies number of training rounds, "batch_size" specifies size of the training round, "learning_rate" controls the weight update rate of the model, and "dropout" reduces overfitting during training. "bilstm_size" is used to specify number of hidden units of a BiLSTM layer.

Table 3. Hyper-parameters of models.

Hyper-Parameters	Parameter Values
max_seq_length	128
epoch	10
batch_size	32
learning_rate	3×10^{-5}
dropout	0.1
bilstm_size	128

4.4. Evaluation Metrics

This paper evaluates the precision (P), recall (R), and F_1 value of all models in NER task of equipment fault diagnosis text, and these three evaluation indicators are widely used in NER tasks [29]. The formulas for the calculation of these three evaluation indicators are given in Equations (12)–(14).

$$P = \frac{TP}{TP + FP} \tag{12}$$

$$R = \frac{TP}{TP + FN} \tag{13}$$

$$F_1 = \frac{2 \times P \times R}{P + R} \tag{14}$$

where *TP* refers to entities identified from texts, *FP* refers to incorrectly identifying non-entities as entities, and *FN* refers to failure to identify real entities [19]. Precision represents the fraction of objects identified as entities by the NER system that are truly entities. Recall represents the proportion of all entities that actually exist that are correctly identified by the NER system. F_1 is the harmonic mean of precision and recall, which is used to comprehensively evaluate the performance of NER systems.

4.5. Experimental Results and Analysis

In this paper, the NER task is implemented for five different models on the equipment fault diagnosis dataset, and the actual performance of models is evaluated on the test set. The results show that the model presented in this paper can effectively extract equipment fault diagnosis entities. Table 4 show examples of extracting entities from equipment fault diagnosis texts. In this section, the experimental results are analyzed in terms of models, hyper-parameters, and entities.

Table 4. The samples of NER in English.

Original Text	Entity Recognition
The transmitter of the search command vehicle has no output of high voltage radiation power.	Equipment: search command vehicle; Part: transmitter; Phenomenon: no output of high voltage radiation power
The auxiliary power supply combined with the cathode current leads to no output of the transmitter high voltage radiated power.	Reason: auxiliary power supply combined cathode current; Part: transmitter; Phenomenon: no output of high voltage radiation power
Replace the auxiliary power supply in the combination of modulator plug-in transmitter high voltage radiation power no output phenomenon eliminated.	Way: Replace the modulator plug-in in the auxiliary power supply combination; Part: transmitter; Phenomenon: no output of high voltage radiation power

4.5.1. Comparison of Different Models

Table 5 shows results of five NER models, and the results are analyzed as follows:

Table 5. Model experimental results.

Model	Precision	Recall	F_1-Value
RoBERTa-wwm-ext-BiLSTM-CRF	0.9457	0.9539	0.9498
BERT-BiLSTM-CRF	0.9347	0.9481	0.9413
RoBERTa-wwm-ext-CRF	0.9259	0.9366	0.9312
BERT-CRF	0.9157	0.9395	0.9275
BiLSTM-CRF	0.8068	0.8251	0.8158

(1) The *P*, *R*, and F_1 values of BiLSTM-CRF and BERT-BiLSTM-CRF reach 0.8068, 0.8251, 0.8158 and 0.9347, 0.9481, 0.9413, respectively. Through comparison, it is found that the introduction of the pre-trained model can effectively improve the *P*, *R*, and F_1 value of NER.

(2) The F_1 values of RoBERTa-wwm-ext-BiLSTM-CRF and RoBERTa-wwm-ext-CRF are 0.9498 and 0.9312, respectively. The results show that the introduction of BiLSTM layer is able to improve the F_1 value, which verifies the effectiveness of adding the BiLSTM layer to the NER task.

(3) The P, R, and F_1 value of RoBERTa-wwm-ext-BiLSTM-CRF and BERT-BiLSTM-CRF are 0.9457, 0.9539, 0.9498 and 0.9374, 0.9481, 0.9413, respectively. The results show that, compared with the basic BERT model, RoBERTa-wwm-ext performs better in NER tasks by introducing dynamic masking technology, discarding NSP tasks, adopting the full-word masking strategy, and increasing the training batch, training step, and training dataset size.

(4) The P, R, and F_1 value of the RoBERTa wwm-ext-BiLSTM-CRF model reach 0.9457, 0.9539, and 0.9498, respectively, which is the best performance among all models, proving that the model presented in this paper has superior performance in the NER task of equipment fault diagnosis.

4.5.2. Effect of Type and Number of Entities

The entities belonging to five types of equipment fault diagnosis, namely equipment name, fault part, fault phenomenon, fault reason, and troubleshooting way, are identified. The recognition results of all models are shown in Table 6, and the experimental results are analyzed as follows.

Table 6. Experimental results of models for the recognition of different entity types and numbers.

Model	Evaluate	Equipment	Part	Phenomena	Reason	Way
RoBERTa-wwm-ext-BiLSTM-CRF	P	0.9855	0.9275	0.8873	0.9296	1.0000
	R	0.9855	0.9275	0.9130	0.9565	0.9859
	F_1	0.9855	0.9275	0.9000	0.9429	0.9929
BERT-BiLSTM-CRF	P	0.9855	0.9412	0.8493	0.9028	1.0000
	R	0.9855	0.9275	0.8986	0.9420	0.9859
	F_1	0.9855	0.9343	0.8732	0.9220	0.9929
RoBERTa-wwm-ext-CRF	P	1.0000	0.9091	0.8219	0.9041	1.0000
	R	1.0000	0.8696	0.8696	0.9565	0.9859
	F_1	1.0000	0.8889	0.8451	0.9296	0.9929
BERT-CRF	P	1.0000	0.8400	0.8571	0.8889	1.0000
	R	1.0000	0.9130	0.8696	0.9275	0.9859
	F_1	1.0000	0.8750	0.8633	0.9078	0.9929
BiLSTM-CRF	P	0.8855	0.8129	0.7333	0.7667	0.8806
	R	0.8855	0.8006	0.7696	0.8420	0.8551
	F_1	0.8855	0.8067	0.7511	0.8028	0.8676

In the recognition of the five types of entities, all models show that the recognition effect of equipment name and troubleshooting way entities are good, and the F_1 score is close to 1. The reason is that the equipment name type entities and the troubleshooting way type entities have relatively fixed formats in the equipment fault diagnosis text, high repeatability in each paragraph of fault text, and are easier to identify than other types of entities. Secondly, the effect of fault part and fault reason entity recognition is also good because these entity types have a large number of labels. In addition, the effect of entity recognition for fault phenomena is relatively low because different personnel will have certain differences in the description and record of fault phenomena, and the difficulty of recognition will be increased.

4.5.3. The Effect of Model Hyper-Parameters

Setting different values for the model hyperparameters sometimes affects the performance of the model, and the most appropriate hyperparameters can be found through experiments. Learning rate and training epochs are important parameters in the model. Learning rate determines the step size of the model parameters update, and affects the training effect and convergence speed of the model. For the adjustment of the model learning rate, the experimental results are shown in Table 7. When the learning rate is set to 3×10^{-5}, the model achieves the best performance.

Table 7. Model performance at different learning rates.

Learning Rate	Precision	Recall	F_1-Value
1×10^{-5}	0.8851	0.8876	0.8863
2×10^{-5}	0.9316	0.9424	0.9370
3×10^{-5}	0.9457	0.9539	0.9498
4×10^{-5}	0.9468	0.9468	0.9468
5×10^{-5}	0.9409	0.9381	0.9395

As the number of iterations grows, Figure 9 illustrates the trend in the F1 score of the model. RoBERTa-wwm-ext-BiLSTM-CRF has lower F_1 values than RoBERTa-wwm-ext-CRF and BERT-CRF in the first two epochs. The F_1 values of all models become stable after the fourth to fifth epoch, among which RoBERTa-wwm-ext-BiLSTM-CRF and BERT-BiLSTM-CRF have the best results, and RoBERTa-wwm-ext-BiLSTM-CRF maintains the largest F_1 value after the third epoch.

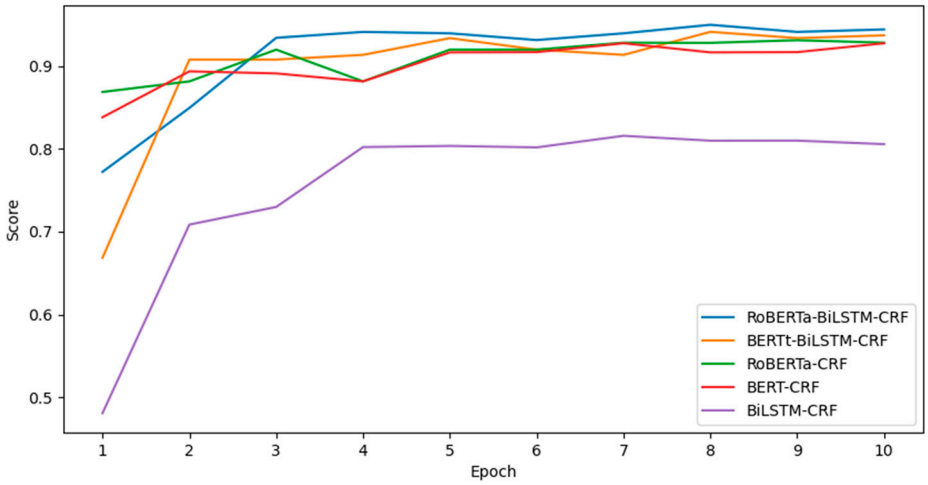


Figure 9. Variation of F1 value as the number of training epochs increases.

4.6. Case Study

The trained RoBERTa-wwm-ext-BiLSTM-CRF equipment fault diagnosis NER model can be directly invoked to recognize the entity of the input text. In order to verify the recognition effect of the model, we conducted a specific case study, and the results are shown in Figure 10.

```
pleae input text: 通信天线车调平支腿工作异常，液压油管破损导致漏油，更换破损油管后故障排除。
['B-equipment', 'I-equipment', 'I-equipment', 'I-equipment', 'I-equipment', 'B-part', 'I-part', 'I-part', 'I-part', 'B-phenomenon', 'I-phenomenon',
  'I-phenomenon', 'I-phenomenon', 'O', 'B-reason', 'I-reason', 'I-reason', 'I-reason', 'I-reason', 'I-reason', 'O', 'O', 'I-reason', 'I-reason', 'O',
  'B-way', 'I-way', 'I-way', 'I-way', 'I-way', 'I-way', 'O', 'O', 'O', 'O', 'O', 'O']
{'通信天线车': 'equipment'}
{'调平支腿': 'part'}
{'工作异常': 'phenomenon'}
{'液压油管破损': 'reason'}
{'更换破损油管': 'way'}
```

Figure 10. Entity recognition of the input text by the model.

In this example, the input text is “the communication antenna car leveling leg is working abnormally, the hydraulic oil pipe is damaged, and the oil leakage is caused, and the problem is solved after replacing the damaged oil pipe”. The model marks the token in the text with their corresponding labels, and then outputs according to the “BIO” labeling

rule. It can be seen that the model recognizes "communication antenna vehicle" in the text as "equipment", that is, "communication antenna vehicle" as the equipment name; "leveling leg" is identified as "part", that is, the "leveling leg" is the fault part; "abnormal work" is identified as "phenomenon", that is, " abnormal work" is the fault phenomenon. "the hydraulic oil pipe is damaged" is identified as "reason", that is, "the hydraulic oil pipe is damaged" is the fault reason. "replace the damaged oil pipe" is identified as "way", that is, "replace the damaged oil pipe" is the troubleshooting way. The recognition results confirm the effectiveness of our proposed model.

5. Conclusions

This paper establishes an NER model for equipment fault diagnosis based on the fusion of RoBERTa-wwm-ext and Deep Learning, aiming to automatically extract named entities from a massive set of equipment fault diagnosis text data. This paper then provides a solid data foundation and support for the construction of equipment fault diagnosis KGs. The model proposed in this paper is used to extract five types of entities from equipment fault diagnosis texts. The average P, R, and F_1 value are 0.9457, 0.9539, and 0.9498, respectively. Based on the experimental results, we draw the following conclusions: (1) After introducing the pre-trained model into NER task, the precision, recall, and F_1 value can be significantly improved. (2) Adding BiLSTM layer can boost model performance. (3) Comparative experiments show that the model proposed in this paper performs well in equipment fault diagnosis NER tasks. (4) When the meaning of entities is clear, the format is fixed, the repeatability is high, and the effect of entity extraction is better.

Although this study is effective for equipment fault diagnosis and named entities recognition, there are still some aspects worthy of improvement and in-depth exploration in future work: (1) For the recognition of fault phenomenon entities, the recognition effect is relatively poor due to the differences in description and record. (2) Due to the small amount of equipment fault diagnosis text data collection and imperfect performance indicators, we plan to build a high-quality equipment fault diagnosis corpus with a larger volume of data and richer entities. (3) The rapid progress of Deep Learning requires us to continuously optimize our model in pursuit of higher performance standards. (4) We plan to carry out related algorithm development work in order to effectively improve the P, R, and F1 values of named entity recognition in the field of equipment fault diagnosis. (5) We plan to apply the model to extract entities from a larger equipment fault diagnosis text and carry out the task of entity relation extraction from fault diagnosis text to build a domain knowledge graph. (6) We will further study the application effect of the model in other fields, such as the structural composition of equipment and other general fields, to further prove the effectiveness of the model.

Author Contributions: Conceptualization, F.G. and L.Z.; methodology, F.G. and B.Z.; software, L.X.; validation, W.W. and W.L.; formal analysis, W.L. and J.Z.; resources, F.G., L.X., and J.Z.; writing—original draft preparation, F.G. and W.L.; writing—review and editing, L.Z. and B.Z.; supervision, W.W.; funding acquisition, B.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported in part by the National Social Science Foundation of China (2022-SKJJ-C-037).

Data Availability Statement: The data that support the findings of this study are available from the corresponding author upon reasonable request.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Ji, S.; Pan, S.; Cambria, E.; Marttinen, P.; Yu, P.S. A Survey on Knowledge Graphs: Representation, Acquisition, and Applications. *IEEE Trans. Neural Netw. Learn. Syst.* **2022**, *33*, 494–514. [CrossRef] [PubMed]
2. Tang, X.; Chi, G.; Cui, L.; Ip, A.W.H.; Yung, K.L.; Xie, X. Exploring Research on the Construction and Application of Knowledge Graphs for Aircraft Fault Diagnosis. *Sensors* **2023**, *23*, 5295–5313. [CrossRef] [PubMed]

3. Deng, J.; Wang, T.; Wang, Z.; Zhou, J.; Cheng, L. Research on Event Logic Knowledge Graph Construction Method of Robot Transmission System Fault Diagnosis. *IEEE Access* **2022**, *10*, 17656–17673. [CrossRef]
4. Liu, L.; Wang, B.; Ma, F.; Zheng, Q.; Yao, L.; Zhang, C.; Mohamed, M.A. A Concurrent Fault Diagnosis Method of Transformer Based on Graph Convolutional Network and Knowledge Graph. *Front. Energy Res.* **2022**, *10*, 837553. [CrossRef]
5. Yu, Y.; Wang, Y.; Mu, J.; Li, W.; Jiao, S.; Wang, Z.; Lv, P.; Zhu, Y. Chinese Mineral Named Entity Recognition Based on BERT Model. *Expert Syst. Appl.* **2022**, *206*, 117727. [CrossRef]
6. Gong, Z.; Cao, Z.; Zhou, S.; Yang, F.; Shuai, C.; Ouyang, X.; Luo, Z. Thermal Fault Detection of High-Voltage Isolating Switches Based on Hybrid Data and BERT. *Arab. J. Sci. Eng.* **2024**, *49*, 6429–6443. [CrossRef]
7. Meng, X.; Jing, B.; Wang, S.; Pan, J.; Huang, Y.; Jiao, X. Fault Knowledge Graph Construction and Platform Development for Aircraft PHM. *Sensors* **2024**, *24*, 231–252. [CrossRef]
8. Yang, X.; Li, H.; Xu, Y.; Shen, N.; He, R. A Text Mining-Based Approach for Comprehensive Understanding of Chinese Railway Operational Equipment Failure Reports. 2024; *preprint*. [CrossRef]
9. Hettne, K.M.; Stierum, R.H.; Schuemie, M.J.; Hendriksen, P.J.M.; Schijvenaars, B.J.A.; van Mulligen, E.M.; Kleinjans, J.; Kors, J.A. A Dictionary to Identify Small Molecules and Drugs in Free Text. *Bioinformatics* **2009**, *25*, 2983–2991. [CrossRef]
10. Chieu, H.L.; Ng, H.T. Named Entity Recognition with a Maximum Entropy Approach. In Proceedings of the Seventh Conference on Natural Language Learning at HLT-NAACL 2003, Edmonton, AB, Canada, 31 May and 1 June 2003; pp. 160–163.
11. Zhao, S. Named Entity Recognition in Biomedical Texts Using an HMM Model. In Proceedings of the International Joint Workshop on Natural Language Processing in Biomedicine and Its Applications (NLPBA/BioNLP), Geneva, Switzerland, 28–29 August 2004; Collier, N., Ruch, P., Nazarenko, A., Eds.; COLING: Geneva, Switzerland, 2004; pp. 87–90.
12. Wang, C.; Ma, X.; Chen, J.; Chen, J. Information Extraction and Knowledge Graph Construction from Geoscience Literature. *Comput. Geosci.* **2018**, *112*, 112–120. [CrossRef]
13. Hochreiter, S.; Schmidhuber, J. Long Short-Term Memory. *Neural Comput.* **1997**, *9*, 1735–1780. [CrossRef]
14. Huang, Z.; Xu, W.; Yu, K. Bidirectional LSTM-CRF Models for Sequence Tagging. *arXiv* **2015**, arXiv:1508.01991.
15. Miao, K.; Han, T.; Yao, Y.; Lu, H.; Chen, P.; Wang, B.; Zhang, J. Application of LSTM for Short Term Fog Forecasting Based on Meteorological Elements. *Neurocomputing* **2020**, *408*, 285–291. [CrossRef]
16. An, Y.; Xia, X.; Chen, X.; Wu, F.-X.; Wang, J. Chinese Clinical Named Entity Recognition via Multi-Head Self-Attention Based BiLSTM-CRF. *Artif. Intell. Med.* **2022**, *127*, 102282. [CrossRef] [PubMed]
17. Devlin, J.; Chang, M.-W.; Lee, K.; Toutanova, K. BERT: Pre-Training of Deep Bidirectional Transformers for Language Understanding. *arXiv* **2018**, arXiv:1810.04805.
18. Guo, Z.X.; Deng, X.L. Intelligent Identification Method of Legal Case Entity Based on BERT-BiLSTM-CRF. *J. Beijing Univ. Posts Telecommun.* **2020**, *44*, 129–134. [CrossRef]
19. Lin, J.; Li, S.; Qin, N.; Ding, S. Entity Recognition of Railway Signal Equipment Fault Information Based on RoBERTa-Wwm and Deep Learning Integration. *Math. Biosci. Eng.* **2023**, *21*, 1228–1248. [CrossRef]
20. Liang, K.; Zhou, B.; Zhang, Y.; He, Y.; Guo, X.; Zhang, B. A Multi-Entity Knowledge Joint Extraction Method of Communication Equipment Faults for Industrial IoT. *Electronics* **2022**, *11*, 979–996. [CrossRef]
21. Kong, Z.; Yue, C.; Shi, Y.; Yu, J.; Xie, C.; Xie, L. Entity Extraction of Electrical Equipment Malfunction Text by a Hybrid Natural Language Processing Algorithm. *IEEE Access* **2021**, *9*, 40216–40226. [CrossRef]
22. Chen, T.; Zhu, J.; Zeng, Z.; Jia, X. Compressor Fault Diagnosis Knowledge: A Benchmark Dataset for Knowledge Extraction From Maintenance Log Sheets Based on Sequence Labeling. *IEEE Access* **2021**, *9*, 59394–59405. [CrossRef]
23. Zhang, Y.; Zhong, Y.; Luo, X. Power Equipment Fault Entity Recognition Based on BERT-CRF Model. In Proceedings of the Fifth International Conference on Artificial Intelligence and Computer Science, Wuhan, China, 26–28 July 2023; pp. 934–941.
24. Zhou, Z.; Zhang, C.; Liang, X.; Liu, H.; Diao, M.; Deng, Y. BERT-Based Dual-Channel Power Equipment Defect Text Assessment Model. *IEEE Access* **2024**, *12*, 134020–134026. [CrossRef]
25. Liu, T.; Liu, Q.; Fu, L. Automation of Book Categorisation Based on Network Centric Quality Management System. *Int. J. Adv. Comput. Sci. Appl. IJACSA* **2024**, *15*, 259. [CrossRef]
26. Zhuang, Y.; Cheng, S.; Kovalchuk, N.; Simmons, M.; Matar, O.K.; Guo, Y.-K.; Arcucci, R. Ensemble Latent Assimilation with Deep Learning Surrogate Model: Application to Drop Interaction in a Microfluidics Device. *Lab Chip* **2022**, *22*, 3187–3202. [CrossRef] [PubMed]
27. Liu, P.; Lv, S. Chinese RoBERTa Distillation For Emotion Classification. *Comput. J.* **2023**, *66*, 3107–3118. [CrossRef]
28. Li, D.; Yan, L.; Yang, J.; Ma, Z. Dependency Syntax Guided BERT-BiLSTM-GAM-CRF for Chinese NER. *Expert Syst. Appl.* **2022**, *196*, 116682. [CrossRef]
29. Liu, X.; Yang, N.; Jiang, Y.; Gu, L.; Shi, X. A Parallel Computing-Based Deep Attention Model for Named Entity Recognition. *J. Supercomput.* **2020**, *76*, 814–830. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.