processes

MDPI

Article

Named Entity Recognition in Mechanical Fault Diagnosis: A Template-Free Prompt Learning Approach

Fenfang Li [1], Ping Luo [1], Junjun Luo [2], Guoyu Qin [2] and Haijun Ma [2,*]

[1] School of Computer Science and Engineering, Northwest Normal University, Lanzhou 730070, China; lifenfang@nwnu.edu.cn (F.L.); luoping@nwnu.edu.cn (P.L.)
[2] Longshou Mine, Jinchuan Group Co., Ltd., Jinchang 737102, China; ljj2005202@163.com (J.L.); qgy1979@163.com (G.Q.)
* Correspondence: 18993500617@163.com; Tel.: +86-18993500617

Abstract: Intelligent prediction, accurate diagnosis, and efficient repair of mechanical equipment faults are critical for ensuring production safety and enhancing efficiency in industrial processes. However, data scarcity and recognition efficiency remain significant challenges in named entity recognition (NER) for mechanical equipment faults. To address these issues, this study proposes a novel NER method based on template-free prompt learning. The model is initialized with limited labeled data and then leverages hidden entity cues in unlabeled data to generate insightful hints, guiding the model through deep retraining. Experimental results demonstrate significant improvements in F1 scores: 29.54%, 22.34%, and 19.67% on the MEFD, CONLL03, and MIT-Movie datasets, respectively, as labeled samples increased from 5 shots to 50 shots. Furthermore, compared to template-based prompt learning methods, the proposed approach achieved F1 score improvements of 9.89%, 12.97%, 9.51%, and 2.21% as labeled samples scaled from 5 shots to 50 shots. The proposed method effectively mitigates data dependency issues, enhances model generalization and application capabilities, and provides robust technical support for the intelligent identification and diagnosis of mechanical equipment faults.

Keywords: industrial production; mechanical equipment malfunction; named entity recognition; prompt learning; generalization ability

check for updates

Academic Editor: Maoyin Chen

Received: 17 March 2025
Revised: 12 April 2025
Accepted: 16 April 2025
Published: 19 April 2025

Citation: Li, F.; Luo, P.; Luo, J.; Qin, G.; Ma, H. Named Entity Recognition in Mechanical Fault Diagnosis: A Template-Free Prompt Learning Approach. *Processes* **2025**, *13*, 1242. https://doi.org/10.3390/pr13041242

Copyright: © 2025 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

In the era of digitalization and intelligent integration, China's manufacturing industry is undergoing a profound transformation. Machinery and equipment, as the mainstays of the manufacturing industry, are leading this change [1]. Under the guidance of the "Made in China 2025" strategy, machinery and equipment are accelerating their evolution towards intelligence, automation, and process orientation. However, intelligent machinery has flaws; wear and tear, fractures, corrosion, and other aging problems remain as performance shackles, which may jeopardize production continuity [2]. With the increasing complexity of equipment and the changing patterns of failures, the traditional labor-intensive maintenance model can no longer meet the high efficiency and low consumption needs of modern manufacturing. Therefore, the integration of the Internet, AI, and big data technology to achieve accurate prediction, intelligent diagnosis, and rapid repair of mechanical equipment failures has become the core path to ensuring production safety, strengthening the competitiveness of the manufacturing industry, and driving the leap towards intelligent manufacturing [3].

Named entity recognition (NER) [4], as a cornerstone technology of Natural Language Processing (NLP), is widely applied in many cutting-edge fields, such as knowledge graph construction [5] and intelligent Q&A systems [6]. Its core function is accurately recognizing and classifying entity objects in text, such as names of people, places, institutions, etc., thereby providing a structured foundation for information processing. In industrial production, especially in machinery and equipment fault diagnosis, NER technology has demonstrated exceptional capability in automatically extracting key data from intricate and complex unstructured maintenance information and constructing a well-organized and informative maintenance guide. This greatly simplifies the maintenance process, enabling intelligent, precise, and rapid responses in equipment management, and significantly improves the efficiency and accuracy of fault diagnosis. However, the unique challenges of NER technology in this field, such as the scarcity of labeled data and high annotation costs, form a critical research topic. Specifically, approaches to effectively utilize the limited data to train a high-performance classification model while avoiding overfitting urgently need to be pursued.

The rise of pre-trained language models has significantly advanced the development of various tasks in the field of NLP, including the recognition of entities in few-shot and zero-shot scenarios. With the widespread adoption of prompt-based learning techniques, the gap between pre-training and downstream tasks has been effectively bridged, leading to breakthroughs, especially in sample-scarce scenarios [7]. To address the challenge of scarce labeled data, template-based NER methods have emerged as a research hotspot [8]. By leveraging predefined rules and templates, such methods demonstrate certain advantages under limited labeled data conditions, effectively mitigating the issue of insufficient data. However, in the context of mechanical equipment fault diagnosis, the scarcity of fault cases, coupled with the complexity and variability of the scenarios, has created a significant bottleneck due to the lack of training data. This limitation hinders the broad application of traditional template-based prompt learning methods in this field. Against that background, exploring more efficient and flexible data utilization strategies to overcome data scarcity has become an important research direction.

To overcome the limitations of prompt learning templates, this paper proposes a NER model based on prompt learning without the need for template construction (template-free prompt NER, TFP-NER). The core idea is to use labeled data as prompts to guide the model in learning the intrinsic features of entities. Specifically, the labeled data are first preprocessed to extract the entities' key features, which are then input into the model as prompts. The model gradually learns the intrinsic features of the entities through these prompts and effectively classifies new entities during testing. To address the challenges faced by NER in low-resource scenarios within the field of mechanical equipment fault diagnosis, this paper proposes a TFP-NER. The main contributions of this paper are as follows:

1. To overcome the limitations of traditional template-dependent prompt learning methods with limited sample data, a template-free prompt NER (TFP-NER) model is proposed, eliminating the need for template construction. This approach reduces template dependency and enhances the model's adaptability and generalization ability across domains, particularly in low-resource domains such as mechanical equipment fault diagnosis.
2. The TFP-NER model leverages the prompt learning technique to effectively recognize new entities even with limited labeled data by mining the implicit representations within the model.
3. Transforming the NER task into a language model (LM) problem further improves the model's performance and flexibility. This approach reduces the dependence on a large amount of labeled data and lowers the cost and difficulty of manual labeling.

2. Related Work

Named entity recognition is a fundamental task in NLP. It has evolved from traditional rule-based and lexicon-based methods to advanced approaches relying on machine learning and deep learning. Each stage of this evolution has had its advantages and limitations, and the main techniques, along with their development, are discussed below.

2.1. Traditional NER Methods

In the early stages, NER primarily relied on rule- and dictionary-based methods. At the 7th IEEE Conference on Applications of Artificial Intelligence, Rau introduced the concept of "noun recognition" for the first time, which laid the foundation for subsequent research [9]. In 1996, information extraction achieved significant results at the 6th Message Understanding Conference (MUC-6), where the concept of named entities was formally proposed for the first time [10]. In 2005, Cafarella et al. [11] extracted named entities from the Web using lexicon-based rules, improving recall without sacrificing precision. Subsequently, Zhang et al. [12] designed rules for syntactic and lexical patterns to filter out non-entities and duplicate entities, resulting in a refined set of entities. As research progressed, rule- and lexicon-based models such as NYU [13], LaSIE-II [14], and NetOwl [15] were proposed.

2.2. Machine Learning-Based NER Methods

Machine learning methods offer superior scalability and adaptability compared to rule-based methods. Chopra et al. [16] applied the Hidden Markov Model (HMM) to NER in Hindi. Still, they achieved poor recognition due to the substantial deviation of the assumptions of chi-squared Markovianity and observation independence from real-world conditions. Sun et al. [17] employed the Conditional Random Fields (CRF) model to recognize biomedical named entities in the literature, integrating shallow syntactic features, boundary detection, and semantic tagging into the CRF model, significantly enhancing its performance.

2.3. Deep Learning-Based NER Methods

In recent years, with the rapid advancement of deep learning technology, deep learning models have been increasingly applied to NLP and have demonstrated an exceptional performance in NER tasks, gradually emerging as the dominant approach in this field. Among these models, Recurrent Neural Networks (RNNs) [18], Convolutional Neural Networks (CNNs) [19], and Transformers [20] are widely used. To address the issue of gradient vanishing that RNNs often encounter when processing long sequences, Hochreiter et al. [21] proposed Long Short-Term Memory (LSTM). Dong et al. [22] introduced the Bidirectional LSTM-CRF (BiLSTM-CRF) framework for NER, setting a new benchmark for Chinese NER. CNNs excel at extracting local features and combining them for richer representations. Kim et al. [23] utilized CNNs to extract semantic features from text, achieving superior results. Subsequently, researchers combined RNNs and CNNs for NER. Ma and Hovy et al. [24] extended this approach to the BiLSTM-CNNs-CRF architecture, where CNNs extract character-level features at the input layer. At the same time, BiLSTMs capture contextual features at the encoding layer. Experiments on the CoNLL2003 dataset demonstrated that this model achieves state-of-the-art performance.

2.4. Pre-Trained Model-Based NER Methods

In the context of the booming development of deep learning, pre-trained models have attracted significant attention in NER tasks. Pre-trained language models can model rich semantic information, thus enabling the resolution of issues such as polysemy. The

BERT [25] model is an outstanding representative of this. Yu et al. [26] employed the BERT model to encode individual characters at the embedding layer. They conducted an NER study on mineral texts, effectively identifying seven types of mineral entities. Zhao et al. [27] conducted an NER study on military domain data, proposing a model named BERT-BiLSTM-CRF, which incorporates a BERT pre-trained language model at the embedding layer to enhance the semantic representation of the text. The experimental results demonstrated a significant improvement in the F1 score.

Most pre-trained language models require substantial data support. To address the issue of data scarcity, few-shot learning [28] and zero-shot learning [29] have been proposed. Few-shot learning enhances the model's ability to generalize to new samples by leveraging limited labeled data. On the other hand, zero-shot learning identifies entities from unseen categories, with common approaches including meta-learning, transfer learning, and reinforcement learning. While these methods have progressed in few-shot and zero-shot tasks, challenges such as model complexity and unstable training persist.

2.5. Prompt Learning-Based NER Methods

Prompt learning (PL), as an emerging approach, reduces the discrepancy between the pre-training phase and the downstream task by designing task prompts to transform the downstream task into a fill-in-the-blank task for language modeling.

Cui et al. [30] first applied prompt learning to NER. Individual prompts are constructed by enumerating all spans in the form of "[X] is an [MASK] entity". The model then categorizes the entity by filling the [MASK] slot. PromptNER is a unified framework for entity localization and type recognition within a prompt learning paradigm. Shen et al. [31] proposed a dual-slot multi-prompt template incorporating location and type slots to enable parallel prediction and significantly reduce inference time, achieving an average 7.7% improvement over prior methods in cross-domain few-shot settings. Ashok and Lipton [32] enhanced PromptNER by integrating entity definitions with few-shot examples and prompting language models to generate interpretable entities, attaining state-of-the-art results on CoNLL, GENIA, and Few-NERD. Layegh et al. [33] introduced ContrastNER, which combines discrete and continuous prompts with contrastive learning, eliminating manual prompt and verbalizer design while achieving a superior performance in low-resource settings. Shao et al. [34] proposed a two-stage prompt learning approach that augments and expands seed spans to produce high-quality candidate entities, followed by type prediction, significantly outperforming existing methods under low-resource conditions.

The proposed NER method based on template-free hint learning transforms the NER task into a language modeling problem by eliminating the reliance on templates and preserving the benefits of hint tuning. The method improves the model's performance and flexibility in low-resource scenarios and reduces the cost and difficulty of manual annotation. Its effectiveness is validated through its application across various domains, providing a new solution for NER in low-resource scenarios.

3. Materials and Methods

The TFP-NER model is a template-based approach for prompt tuning. It utilizes the mapping relationship between answers and labels in prompt learning to effectively exploit the Prompt Learning Model (PLM) capabilities in named entity recognition tasks.

3.1. Model Structure

Figure 1 shows the model structure. The model reformulates the NER task as a language modeling (LM) problem by generating tag-word vectors from the dataset and uti-

lizing the LM embedding prototypes without any template. Overall, the model abandons the template construction process while maintaining the word prediction paradigm of the pre-trained model to predict class-specific tag words at entity positions.

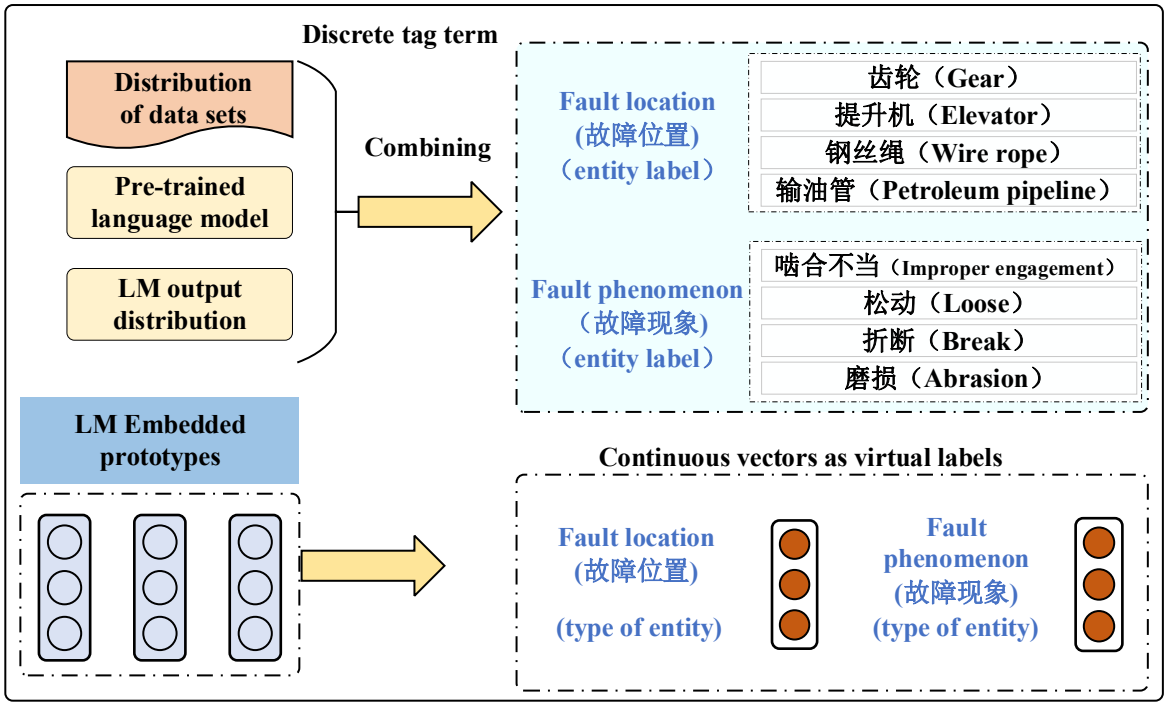


Figure 1. TFP-NER model structural diagram.

3.2. NER Based on Template-Free Prompt Learning

Before performing NER tasks based on template prompt learning, we need to manually define positive and negative sample templates to establish classification criteria. As shown in Table 1, the classification criteria are divided into four categories: In text classification, positive samples are texts belonging to the target category, while negative samples are texts from other or interfering categories. In named entity recognition (NER), positive samples are text segments of the target entity type, while negative samples are non-target entities or segments with incorrect boundaries. In text similarity matching, positive samples are sentence pairs with high semantic similarity, while negative samples are semantically unrelated sentence pairs. In multi-label classification, positive samples are texts containing the target label, while negative samples are texts that do not contain the target label or contain incorrect labels.

The NER method based on template cue learning differs from traditional approaches that rely on sequence annotation. It aims to study low-resource scenarios and treats NER as a complementary cueing task. Positive and negative sample templates must be manually defined in the recognition process. For each phrase in a given sentence, if it is an entity, the positive sample template $< X_(i:j) >$ is $< L_k >$ entity is used; if it is not an entity, the negative sample template $< X_(i:j) >$ is not an entity is used.

An example of the NER process based on template prompt learning is shown in Figure 2. For example, for the input text "The wirerope has broken ("钢丝绳出现断裂")", several sets of template text corresponding to whether each phrase is a specific entity type must be constructed, such as "broken is a failure phenomenon". In the training phase, template samples of all entity and non-entity types are built according to the labels. At the same time, multiple sets of text generated from the templates are predicted in the decoder. In the case of "The wirerope has broken ("钢丝绳出现断裂")", the decoder aims to create the text

"broken is a failure phenomenon entity" using the template output. The decoder can score the text constructed from the template based on the original input. Suppose the sentence "The broken belongs to the entity of the failure phenomenon ("断裂属于故障现象实体") (FP)" scores highly. In Figure 2, "×" indicates that the score is too low and will not be used as a classification result, while "✓" indicates that the score is the highest and will be used as a classification result. In that case, it indicates that the sentence is correct, and "broken" can be extracted as an entity corresponding to the failure phenomenon.

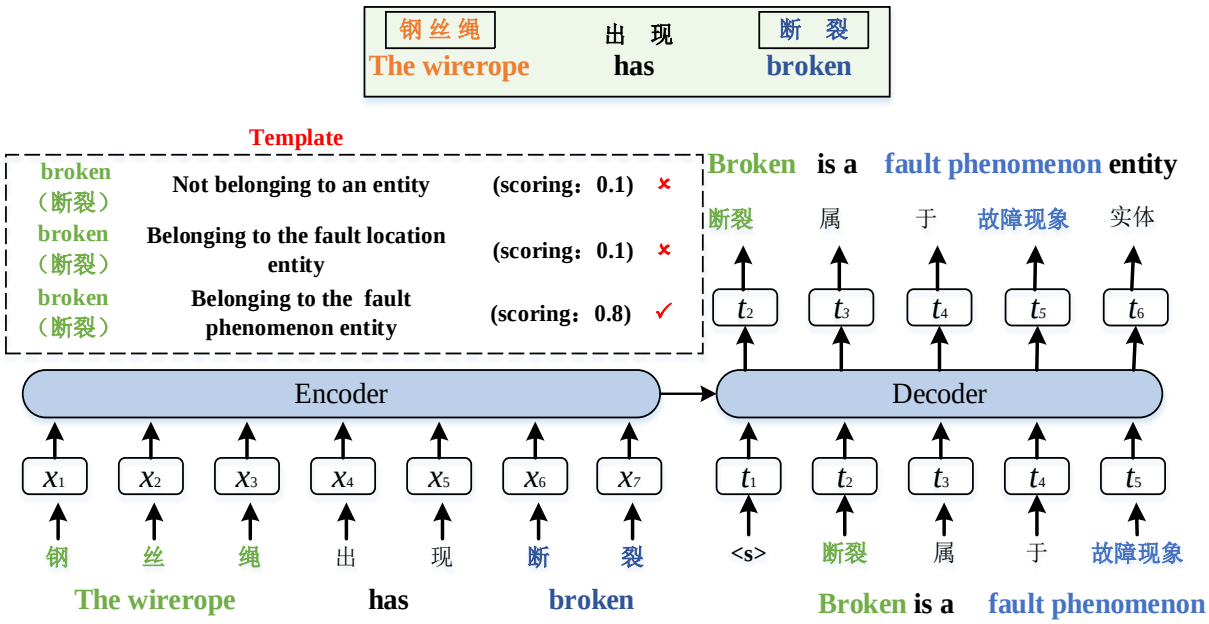
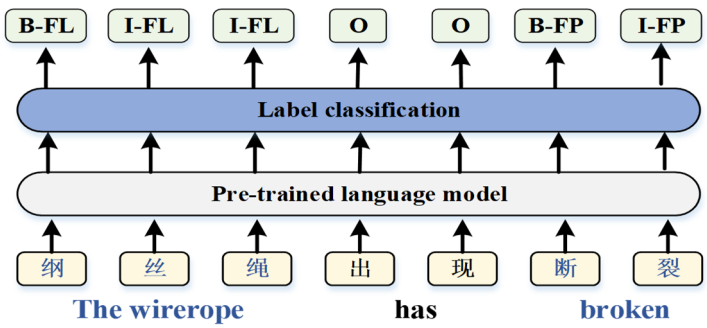


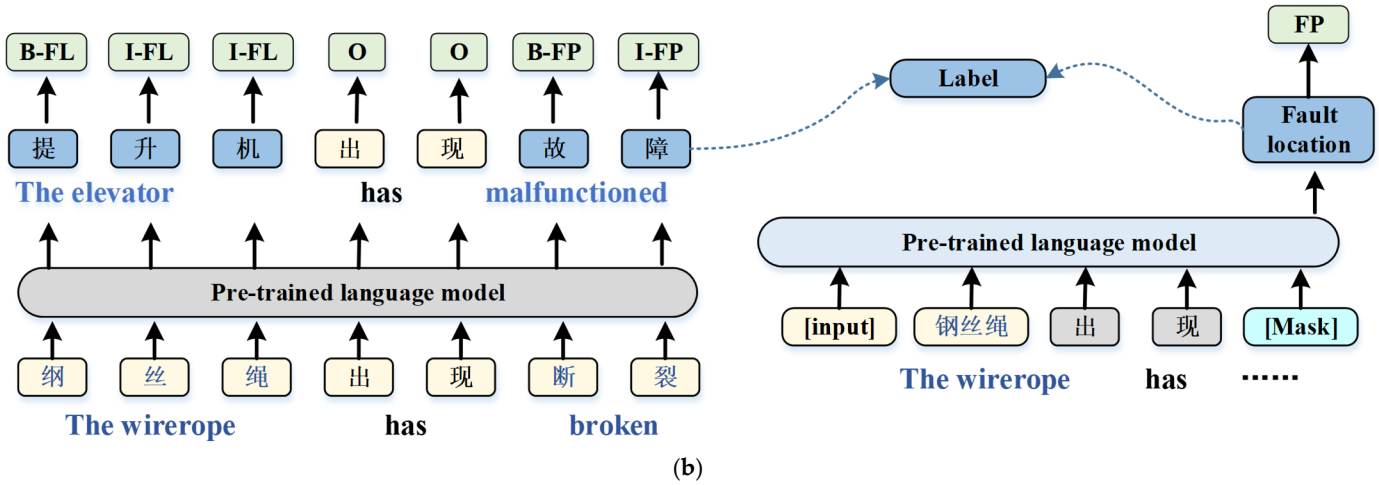
Figure 2. NER process based on template prompt learning.

To solve the efficiency problem of template-based prompt tuning, the TFP-NER model constructs different templates for entity types and candidate entities for input comparisons. This approach can be accomplished with only one forward computation during the entity recognition process, effectively reducing redundant computations and improving the algorithm's efficiency. As shown in Figure 3, the overall form of the TFP-NER model is consistent with the traditional sequence labeling task paradigm. However, the difference lies in the output: the conventional sequence annotation task outputs the entity type label corresponding to each character in the input sentence, whereas the TFP-NER model selects a representative character to denote the position of the entity. For non-entity parts, the original word is retained.



(**a**)

Figure 3. *Cont.*



(**b**)

Figure 3. Comparison of different fine-tuning methods for NER. (**a**) Standard fine-tuning method. (**b**) TFP-NER method.

As shown in Figure 3b, when the input "The wirerope has broken ("钢丝绳出现断裂")" is given, TFP-NER is trained to predict a label word "elevator ("提升机")" at the location of the entity "wirerope ("钢丝绳")" as an indication of the label "FL". For the non-entity word "has," the model still predicts the original word. Firstly, we construct a set of labeled words $\mathcal{V}_l$, which is connected to the set of task labels through the mapping function $M : \mathbb{Y} \rightarrow \mathcal{V}_\iota$. Given an input sentence X, which contains a series of elements $\{x_1, \ldots, x_n\}$, and a corresponding tag sequence Y, the sequence contains the tags $\{y_1, \ldots, y_n\}$. Construct a target sentence $\mathrm{X}^{\mathrm{TFP}} = \left\{x_1, \ldots, M_{(y_i)}, \ldots, x_n\right\}$ by replacing y_i at entity position i (assuming y_i is an entity label) with the corresponding labeled word $M_{(y_i)}$, and keeping the original word entity position. Then, given the original input X, train the LM to maximize the probability $\mathrm{P}\left(X^{TFP}|X\right)$ of the target sentence $\mathrm{X}^{\mathrm{TFP}}$, which is expressed as Equation (1):

$$L_{TFP} = -\sum_{i=1}^{n} log\ P\left(x_i = x_i^{TFP}\middle|X\right) \tag{1}$$

$\mathrm{P}\left(x_i = x_i^{TFP}|X\right) = \mathrm{Softmax}(W_{lm} \cdot h_i)$ in Equation (1). When W_{lm} is a parameter of the head of the pre-trained language model, the approach aligns with the goal of TFP-NER by reusing the entire pre-trained model and avoiding the introduction of new parameters during fine-tuning. This strategy not only eliminates the complexity of template construction in the NER task but also preserves the prompt learning method's strong mapping ability.

During the testing process, the test sentence X is directly inputted into the model, and the ith text is labeled as a probabilistic modeling of the class $y \in \gamma$, denoted as Equation (2).

$$P(y_i = y|X) = P(x_i = M(y)|X) \tag{2}$$

Only one decoding is needed in this process to obtain all the labels of each sentence, which is significantly more efficient than a template-based prompt query. This approach addresses the efficiency problem of hint-based NER learning. However, it also introduces a new challenge: selecting the labeling characters that represent the entity types.

3.3. Labeling Word Selection

Object guidance is a strategy or mechanism in the TFP-NER methodology that guides the model's training process based on the objects or concepts of named entity classes. It aims to select one or more label words for each named entity class. These label words play a key role in the model's training process as representations of their respective entity classes.

The model learns to associate these label words with the corresponding entity categories, enabling it to accurately assign the correct labels to entities in the text during prediction.

Label word selection can be inconsistent when labeled samples are scarce due to sampling randomness. Leveraging unannotated data and dictionary-based annotations can improve the consistency and effectiveness of label word selection. The TFP-NER model explores four distinct label word search strategies in depth, aiming to enhance the accuracy and efficiency of the NER task.

3.3.1. Data Distribution Search

The most frequently occurring words for each category are identified from the dataset and used as tag words. The most intuitive method of Data Distribution Search (Data Search) is to select the words with the highest frequency of occurrence in a given class in the corpus. Specifically, when searching for tagged words in class C, the frequency $\emptyset(x = w, y^* = C)$ is calculated for each word $w \in V$ labeled as C, and the word with the highest frequency of occurrence is selected by sorting it as shown in Equation (3):

$$M(C) = argmax_w \emptyset(x = w, y^* = C) \tag{3}$$

3.3.2. Tag Term Selection Using Pre-Trained Language Models

To perform tag word selection using pre-trained language models, first, the input samples (X, Y^*) are fed into the LM, which outputs the probability distribution $P(Xi = w|X)$ that predicts each word $w \in V$ at each position. Secondly, an indicator function $Itopk(Xi = w|X) \rightarrow \{0, 1\}$ is introduced, which is used to determine whether a word w belongs to the k predictions at position i in the sample. The labeled words of class C can be obtained by Equation (4).

$$M(C) = argmax_w \sum_{(X,Y^*)\in\mathcal{D}} \sum_i^{|X|} \emptyset_{topk}(\hat{x}_i = w, {y_i}^* = C) \tag{4}$$

where $\emptyset_{topk}(x_i = w, {y_i}^* = C) = I_{topk}(x_i = w|X, Y^*)$. $I({y_i}^* = C)$ denotes the frequency of the occurrence of w in the first k predictions of the location labeled class C.

3.3.3. Tag Term Selection for Fusing Data Distribution Search and Language Model Search

This paper proposes a method that combines the data distribution search and language model search to more comprehensively consider the data characteristics and the language model's capability in selecting accurate and contextualized tag terms. This approach enables the selection of tag terms to account for the actual data's tendencies and the language model's deep contextual understanding. Additionally, the method incorporates a threshold filtering mechanism to prevent the repetition of tag terms across different categories. Specifically, the label words for category C can be derived using Equation (5).

$$M(C) = argmax_w \{\sum_{(X,Y^*)\in\mathcal{D}} \sum_i^{|X|} \emptyset_{topk}(x_i = w, {y_i}^* = C) \cdot \sum_{(X,Y^*)\in\mathcal{D}} \sum_i^{|X|} \emptyset_{topk}(\hat{x}_i = w, {y_i}^* = C)\} \tag{5}$$

3.3.4. Using Virtual Vectors as Labeling Words

To capture the semantic information of the text more flexibly, virtual vectors are introduced as labeled words. These vectors are not limited to predefined vocabularies but can exist as continuous representations. A more intuitive approach is to adopt Prototypical Networks, which use the average vector of word embeddings to define prototypes for the words in each class. However, since averaging the embeddings of all words in a

class is computationally expensive, we propose using the average vector of the top-k high-frequency words selected by the previous method, as demonstrated in Equation (6).

$$M(C) = \frac{1}{|\mathcal{V}_C|} \sum_{w \in \mathcal{V}_C} f_{\varnothing}(w) \tag{6}$$

where V_C is the set of labeled words obtained by finding the *top-k* words corresponding to Equations (3)–(5), and $f_{\varnothing}(\cdot)$ denotes the embedding function of the pre-trained model.

4. Experimental Results and Analysis

4.1. Experimental Environment and Parameter Settings

The experiments in this paper are conducted on the Ubuntu operating system as the experimental platform, utilizing a high-performance NVIDIA Tesla P100 GPU to accelerate computations. We employ Python 3.7 to set up the environment based on the PyTorch 1.8.1 framework, integrate the BERT pre-trained language model as the underlying encoder, and use the Adam optimizer to optimize the model parameters adaptively. The model is considered to have reached its optimal state under the current experimental conditions when it achieves the highest F1 score on the validation set. This criterion is used to determine the final parameter values. The detailed experimental parameter settings are provided in Table 1.

Table 1. Experimental parameter settings.

Parameters	Values
Dimension of the word vector	200
Learning rate	1×10^{-5}
Batch size	16
Dropout	0.1
Hidden layer	768
Number of attention heads	12
Maximum sentence length	128
Maximum fused vocabulary information per word	3

4.2. Experimental Data and Evaluation Index

This paper validates the model performance on the self-constructed Mine Elevator Fault Diagnosis Dataset (MEFD). Additionally, two publicly available datasets, CONLL03 [35] (Conference on Computational Natural Language Learning, https://www.cnts.ua.ac.be/conll2002/ner/) (accessed on 15 April 2025) and MIT-Movie [36] (https://tianchi.aliyun.com/dataset/145106/) (accessed on 15 April 2025), are used to evaluate the model's generalization capabilities. Experiments use K ∈ {5, 10, 20, 50}. Three training sets are selected for each K-shot experiment, and the experiment is repeated four times on each set. Precision (P), Recall (R), and F1 score (F1) are used as the evaluation metrics for the classification task.

(a) The MEFD dataset is a high-quality corpus of fault-named entities manually constructed by the research group. It is built by collecting point inspection forms and maintenance logs from the lifting work area of a state-owned mega non-ferrous metal group, fault-related journal papers, and master's and doctoral theses in the field. The dataset contains a total of 7733 entities. Using the Label-Studio tool, the entities are annotated, and the distribution of different entity types in the MEFD dataset is presented in Table 2; other categories of entities are represented in Figure 4.

Table 2. MEFD named entity annotation system and quantity.

Abbreviation	Entity Category	Start of Entity	Within Entity	Example	Quantity	Ratio
FLN	Fault Location (故障位置)	B-FLN	I-FLN	Screws, oil pumps, wire rope, hoist drums (螺钉、油泵、钢丝绳、提升机卷筒)	3749	48.5%
FET	Fault Effects (故障影响)	B-FET	I-FET	Alarm light flashing, lighted alarm (报警灯闪烁、亮灯报警)	110	1.4%
RMS	Repair Measures (修护措施)	B-RMS	I-RMS	Check reconnection, optimize lubrication station cooling system (检查重新连接、优化润滑站冷却系统)	801	10.4%
FCE	Fault Cause (故障原因)	B-FCE	I-FCE	High oil temperature, overuse (油温过高、超期使用)	791	10.2%
FPN	Failure Phenomenon (故障现象)	B-FPN	I-FPN	Improper engagement, wrong winding connection (啮合不当、绕组连接错误)	2282	29.5%

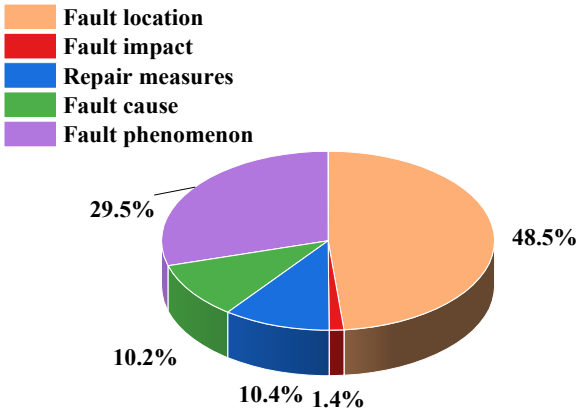


Figure 4. Proportion of different types of entities in the MEFD dataset.

An example of entity labeling is illustrated in Figure 5. The original text shown in Figure 5 is "齿轮啮合不当导致油温过高 ("Improper gear meshing leads to high oil temperature")". In this example, there are three entities: "gear", "improper meshing", and "high oil temperature". "齿轮 ("Gear")" is the fault location (FL) entity, "啮合不当 ("improper meshing")" is the fault cause (FC) entity, "油温过高 ("high oil temperature")" is the fault phenomenon (FP) entity, and "导致("leads to")" is not an entity. The corresponding entities were annotated using the "BIO" annotation method, with the first token labeled as "B-" and the remaining tokens labeled as "I-".

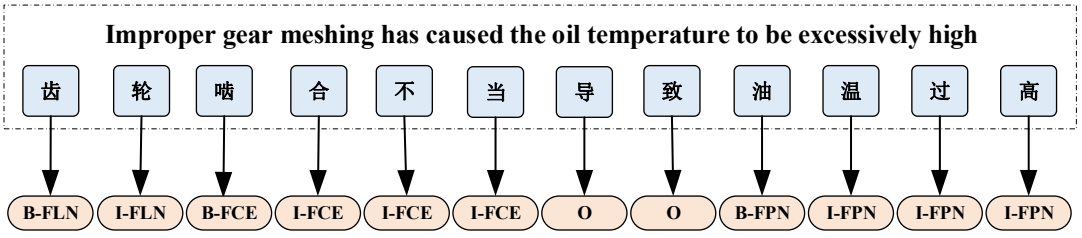


Figure 5. Example of "BIO" annotation in MEFD dataset.

(b) The CONLL03 dataset is a benchmark dataset for evaluating the performance of various NLP algorithms in entity recognition and relation extraction. It consists of English news articles containing labeled entities and relations, with the specific types shown in Table 3. The dataset includes 1071 sentences in the training set and 288 in the test set.

Table 3. Entity types and numbers in the CONLL03 dataset.

Type of Entity	Number of Entities
PER	1685
LOG	1968
ORG	978
Other	705

(c) The MIT-Movie dataset is used for movie script parsing, which mainly focuses on the structured elements of the script. The MIT-Movie dataset contains 12 entity types in scripts: Title, Viewers' Rating, Year, Genre, Director, MPAA Rating, Plot, Actor, Trailer, Song, Review, and Character. The training set size is 8.8 k, and the test set is 2.4 k.

4.3. Analysis of Experimental Results

In this section, the performance of the TFP-NER model is comprehensively evaluated. First, the optimal baseline model of the day is compared to verify its superiority in terms of overall performance. Subsequently, further ablation experiments are designed to validate the effectiveness of each component in the model, thereby providing insights into the key factors for model performance improvement.

4.3.1. Baseline Model Experiments

(1) Effect of K-shots labeled data on the MEFD dataset

To verify the TFP-NER model's entity recognition performance, the BERT-tagger [25], NNShot model [37], StructShot model [37], and TemplateNER [30] models were selected for comparative experiments on the MEFD dataset. The results are presented in Table 4 and Figure 6.

Table 4. The impact of K-shots annotated data on the MEFD dataset (%).

K-Shots	K = 5			K = 10			K = 20			K = 50		
	P	R	F1	P	R	F1	P	R	F1	P	R	F1
BERT-tagger	38.89	38.26	38.45	42.56	42.32	42.4	44.18	43.56	43.68	49.35	49.56	49.37
NNShot	42.98	41.65	42.31	59.63	58.98	59.24	63.76	62.16	62.55	66.38	67.54	66.89
StructShot	44.63	46.63	45.82	62.33	62.36	62.37	63.54	62.45	62.79	68.71	69.89	69.51
TemplateNER	43.01	43.15	43.04	57.63	58.12	57.86	62.27	61.74	61.52	66.47	66.23	66.38
TFP-NER	51.23	51.46	51.32	66.90	66.78	66.86	69.45	68.74	68.92	80.63	80.89	80.86

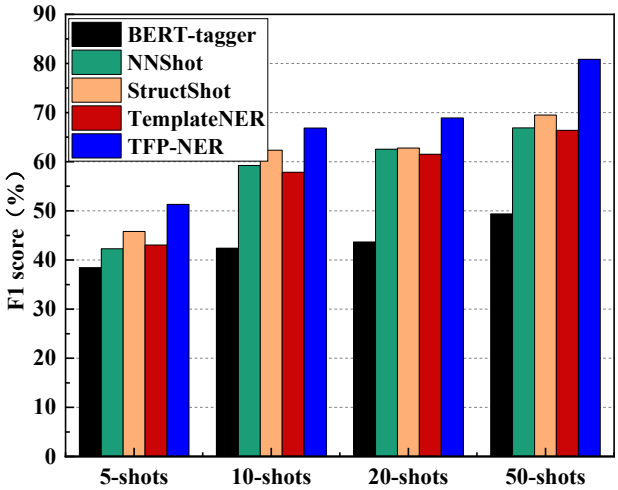


Figure 6. The impact of K-shots annotated data on the MEFD dataset.

From the experimental results presented in Table 4 and Figure 6, it is evident that on the MEFD dataset, the F1 scores of all models significantly improve as the value of K increases from 5 to 50. The TFP-NER model substantially outperforms all baseline models and demonstrates a marked superiority over the template-based TemplateNER model, proving its effectiveness compared to traditional template methods. In contrast, the BERT-tagger model underperforms in the few-shot learning environment, highlighting its limitations in this domain. Additionally, despite the theoretical potential of the NNShot model as a metric-based learning approach, it fails to surpass BERT-tagger in real-world scenarios with limited data resources, revealing the constraints of such models in low-resource settings. On the other hand, the StructShot model remains a competitive baseline despite incorporating lexicons, unlabeled data, and a structural decoder. However, the TFP-NER model, with its adoption of the Viterbi decoder, consistently demonstrates a superior performance, further solidifying its leading position.

(2) Impact of K-shots labeled data on the CoNLL03 dataset

The CoNLL03 dataset, a classic public dataset in the NER domain, encompasses many entity types, including names of people, places, organizations, and institutions. Consequently, the CoNLL03 dataset was selected as the experimental benchmark to comprehensively evaluate the accuracy and stability of the TFP-NER model in entity recognition tasks. As illustrated in Table 5, the same experimental settings and evaluation metrics were applied in the comparative experiments.

Table 5. The impact of K-shots annotated data on the CoNLL03 dataset (%).

K-Shots	K = 5			K = 10			K = 20			K = 50		
	P	R	F1	P	R	F1	P	R	F1	P	R	F1
BERT-tagger	41.35	42.23	41.87	58.65	60.23	59.91	69.37	68.51	68.66	72.65	74.32	73.20
NNShot	42.58	42.21	42.31	58.79	60.32	59.24	67.52	66.74	66.89	71.36	72.88	72.63
StructShot	45.63	46.96	45.82	61.63	63.35	62.37	68.45	70.23	69.51	74.89	74.36	74.73
TemplateNER	43.62	42.63	43.04	58.36	57.54	57.86	67.32	66.57	66.38	72.68	72.76	72.71
TFP-NER	53.68	51.17	51.32	66.45	67.54	66.86	71.65	70.89	71.23	72.89	73.10	73.66

The experimental results demonstrate that the TFP-NER model significantly enhances performance on the CoNLL03 dataset. At K = 5, the F1 score reaches 51.32%, approximately 10% higher than other models, indicating a more pronounced performance improvement. As K increases from 5 to 50, the F1 score and other metrics show significant enhancement across different models. Specifically, at K = 50, the F1 score achieves 73.66%. These findings suggest that the K-shot training method is highly effective in accurately predicting new and unseen categories. Moreover, the TFP-NER model can capture entity boundaries and semantic information in text, effectively recognizing various complex entity types and demonstrating a robust generalization ability.

(3) Impact of K-shots labeled data on MIT-Movie dataset

The MIT-Movie dataset, designed for movie script parsing, primarily focuses on scripts' structured elements. It includes 12 types of entities, such as actor names and movie titles. Due to its comprehensive nature, the MIT-Movie dataset was selected as the experimental benchmark to rigorously evaluate the accuracy and stability of the TFP-NER model in entity recognition on public datasets. The experimental results are presented in Table 6.

Table 6. The impact of K-shots annotated data on the MIT-Movie dataset (%).

K-Shots	K = 5			K = 10			K = 20			K = 50		
	P	R	F1	P	R	F1	P	R	F1	P	R	F1
BERT-tagger	37.94	37.76	37.77	50.23	49.61	49.85	58.67	57.31	57.89	71.35	70.58	70.72
NNShot	39.56	38.54	38.76	51.34	49.76	50.42	59.44	58.72	58.63	71.55	70.32	70.43
StructShot	41.57	40.96	41.25	51.88	51.24	51.39	59.45	58.88	59.07	71.77	71.02	70.58
TemplateNER	44.76	44.21	44.34	48.16	47.53	47.58	58.48	56.57	57.62	72.54	71.53	71.69
TFP-NER	55.18	53.64	54.23	61.23	59.78	60.55	67.78	66.92	67.13	74.55	73.67	73.90

The experimental results in Figure 6 demonstrate that the TFP-NER model achieves a significant performance improvement on the MIT-Movie dataset, with an F1 score of 54.23% at K = 5, substantially outperforming the other baseline models. When K is increased to 50, its F1 score rises to 73.9%, surpassing the performance of the different models by approximately 3%. These results indicate that the template-free named entity recognition method proposed in this study effectively captures entity boundaries and semantic information in the text while exhibiting superior recognition capabilities for more complex entities.

4.3.2. Influence of Tag Word Selection on Recognition Effect

As introduced in Section 3.3 regarding tag word selection methods, this section conducts ablation experiments on four tag word selection methods based on TFP-NER. These methods include the data distribution search (Data Search), language model search (LM Search), data distribution search fused with a language model (Data + LM Search), data distribution search fused with virtual word selection (Data + Virtual), language model fusion with virtual word selection (LM + Virtual), and data distribution search fusion with a language model and virtual word selection (Data + LM + Virtual), totaling six methods. The experimental results are presented in Figure 7.

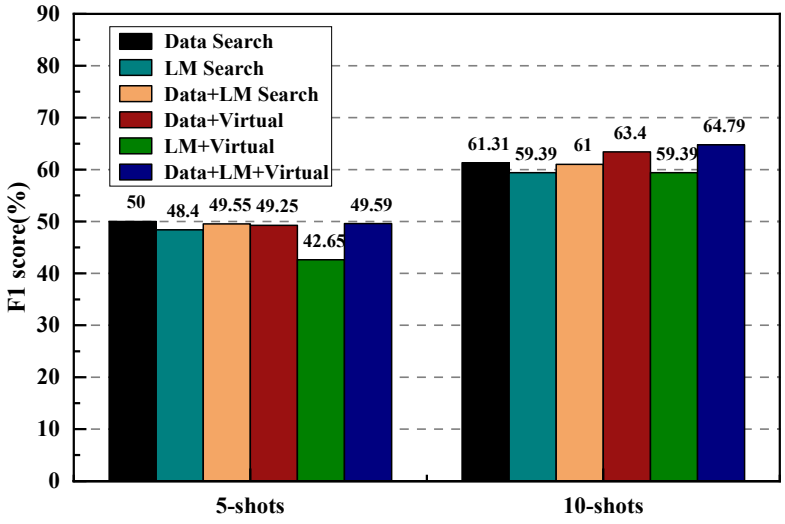


Figure 7. The impact of label word selection methods on TFP-NER.

As shown in Figure 7, virtual word selection methods generally demonstrate superiority over data distribution search word selection methods. Among these, the approach that combines the data distribution search with a language model (LM) distribution to select high-frequency words stands out. Its advantage likely stems from the fact that this method simultaneously captures the a priori data features of the target dataset and the rich contextual information provided by the pre-trained language model (PLM). The fusion of these

two types of information effectively contributes to performance improvement. However, it is worth noting that in a few-shot learning scenario with (K = 5), relying solely on the distribution of the pre-trained language model for a search often fails to achieve the desired results. This reveals that while the pre-trained model contains a broad range of generalized knowledge, its adaptability may be limited when applied to specific tasks, leading to suboptimal performance.

4.3.3. Lexicon Sampling Experiments

The TFP-NER model integrates untagged data and dictionary annotations for tagged word selection. In this experiment, the strengths and limitations of the TFP-NER method in practical applications are further revealed by examining the impact of dictionary quality on the performance of the MEFD dataset. To comprehensively assess the effects of lexicon quality, an entity frequency weight sampling approach was adopted to generate lexicons of varying sizes, ranging from 5% to 80% of the original lexicon size. The impact of TFP-NER and the baseline model on lexicon size is illustrated in Figure 8.

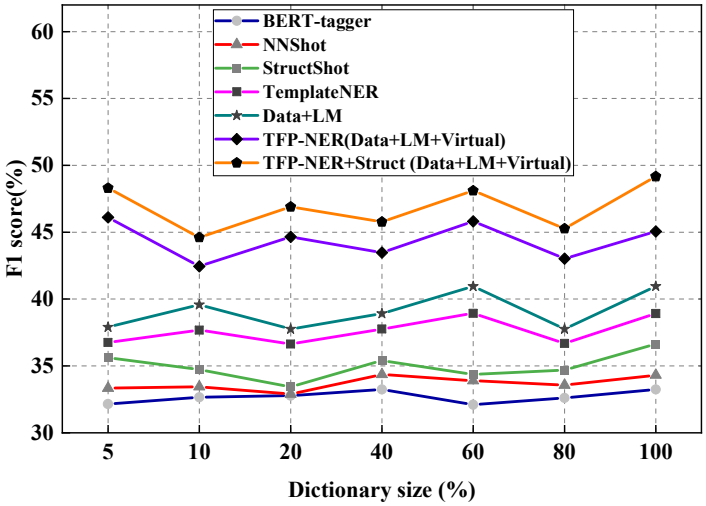


Figure 8. Dictionary sampling experiment.

Figure 8 demonstrates that the F1 scores of different models generally increase as the proportion of the dictionary grows. The TFP-NER model maintains a high performance even under extreme conditions where the dictionary vocabulary size is only 5%, utilizing the Data & LM + Virtual selection strategy. This fully highlights the model's powerful learning capability. In contrast, the TFP-NER model, which relies solely on the Data & LM strategy, performs significantly worse in terms of lexical quality. This discrepancy may arise because the Data & LM approach fails to fully leverage virtual annotation data to enhance the model's generalization ability. Its performance exhibits considerable fluctuations when faced with lexicons of varying quality, revealing its limitations in lexicon dependency.

In summary, this experiment further analyzes the advantages and potential limitations of the TFP-NER model by investigating the effect of dictionary quality on its performance. The results demonstrate that the TFP-NER model can effectively recognize fewer samples and maintain a stable performance across different dictionary qualities. This important finding provides valuable guidance for selecting dictionaries in practical applications and identifies directions for further optimization of the TFP-NER model.

5. Conclusions

This study proposes a named entity recognition (NER) method based on template-free prompt learning to enhance text recognition capabilities in mechanical equipment fault diagnosis. By leveraging prompt information to guide the model's learning process, the proposed method improves the identification and understanding of key details in fault descriptions. The experimental results demonstrate that the method achieves a superior performance on the mechanical equipment fault dataset, significantly enhancing the accuracy and efficiency of fault diagnosis through the construction of instructive prompts. Comparative experiments on a public dataset and ablation studies further validate the method's effectiveness. Future work will focus on optimizing the prompt generation algorithm to improve prompt accuracy and effectiveness, enabling the model to better understand and process complex text descriptions. Additionally, we plan to extend the application of prompt learning and NER methods to other low-resource domains, such as healthcare and agriculture, to support intelligent development across a wider range of fields.

Despite the excellent performance of the template-free prompt learning-based named entity recognition method (TFP-NER) in machinery fault diagnosis, its flexibility, ability to quickly respond to new requirements, and reduction in cross-domain development costs make it highly feasible for adoption across different industries. However, it currently faces the challenge of a large workload associated with manually defining templates. This issue calls for the development of complementary automated tools to reduce the manual effort required for template creation. Future work will focus on enhancing the performance and applicability of the TFP-NER model by introducing datasets from more domains for testing, ensuring its ability to effectively handle complex terminology and specialized knowledge. Through these improvements, we aim to further enhance the model's flexibility and intelligence.

Author Contributions: Conceptualization, F.L.; methodology, F.L.; software, P.L.; validation, J.L. and G.Q.; formal analysis, P.L.; investigation, H.M.; resources, F.L.; data curation, J.L.; writing—original draft preparation, F.L.; writing—review and editing, G.Q.; visualization, H.M.; supervision, G.Q. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China (Grant 62162056) and the Natural Science Foundation of Gansu Province (24JRRA127).

Data Availability Statement: The data presented in this study are available on request from the corresponding author due to privacy reasons.

Conflicts of Interest: Authors Junjun Luo, Guoyu Qin and Haijun Ma were employed by the company Longshou Mine, Jinchuan Group Co., Ltd. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

1. Lei, Y.G.; Jia, F.; Kong, D.T.; Lin, J. Opportunities and challenges of intelligent fault diagnosis of machinery under big data. *J. Mech. Eng.* **2018**, *54*, 94–104. [CrossRef]
2. Li, C. Exploration of safety management problems and countermeasures of mining electromechanical machinery and equipment. *China Met. Bull.* **2023**, 80–82.
3. Wu, N. Design and Implementation of Fault Diagnosis System Based on Deep Learning. Master's Thesis, University of Chinese Academy of Sciences (Shenyang Institute of Computing Technology, Chinese Academy of Sciences), Shenyang, China, 2021.
4. Basra, J.; Saravanan, R.; Rahul, A. A survey on named entity recognition—Datasets, tools, and methodologies. *Nat. Lang. Process. J.* **2023**, *3*, 100017.
5. 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.* **2020**, *33*, 494–514. [CrossRef] [PubMed]

6. Ansari, S.; Raut, R.; Shah, K.B. A review of question answering systems: Approaches, challenges, and applications. *Int. J. Comput.* **2023**, *46*, 25–33.
7. Zhang, Q.T.; Wang, Y.C.; Wang, H.X.; Wang, J.X.; Chen, H. A research review on fine-tuning techniques for large language models. *Comput. Eng. Appl.* **2024**, *60*, 17–33.
8. Li, H.P.; Ma, B.; Yang, Y.T.; Wang, L.; Wang, Z.; Li, X. A chapter-level event extraction method based on slot semantic augmented cue learning. *Comput. Eng.* **2023**, *49*, 23–31.
9. Rau, L.F. Extracting company names from text. In Proceedings of the Seventh IEEE Conference on Artificial Intelligence Application, Miami Beach, FL, USA, 24–28 February 1991; pp. 29–32.
10. Grishman, R.; Sundheim, B. Message understanding conference-6: A brief history. In Proceedings of the 16th International Conference on Computational Linguistics, Copenhagen, Denmark, 5–9 August 1996; pp. 466–471.
11. Etzioni, O.; Cafarella, M.; Downey, D.; Popescu, A.-M.; Shaked, T.; Soderland, S.; Weld, D.S.; Yates, A. Unsupervised named-entity extraction from the Web: An experimental study. *Artif. Intell.* **2005**, *165*, 91–134. [CrossRef]
12. Zhang, S.; Elhadad, N. Unsupervised biomedical named entity recognition: Experiments with clinical and biological texts. *J. Biomed. Inform.* **2013**, *46*, 1088–1089. [CrossRef] [PubMed]
13. Grishman, R. The NYU system for MUC-6 or where's the syntax? In Proceedings of the Sixth Message Understanding Conference (MUC-6), Columbia, MD, USA, 6–8 November 1995.
14. Humphreys, K.; Gaizauskas, R.; Azzam, S.; Huyck, C.; Mitchell, B.; Cunningham, H.; Wilks, Y. University of Sheffield: Description of the LaSIE-II system as used for MUC-7. In Proceedings of the Seventh Message Understanding Conference (MUC-7), Fairfax, VA, USA, 29 April–1 May 1998.
15. Black, W.J.; Rinaldi, F.; Mowatt, D. FACILE: Description of the NE system used for MUC-7. In Proceedings of the Seventh Message Understanding Conference (MUC-7), Fairfax, VA, USA, 29 April–1 May 1998.
16. Chopra, D.; Joshi, N.; Mathur, I. Named entity recognition in Hindi using hidden Markov model. In Proceedings of the 2016 Second International Conference on Computational Intelligence & Communication Technology (CICT), Ghaziabad, India, 12–13 February 2016; IEEE: New York, NY, USA, 2016; pp. 581–586.
17. Sun, C.; Guan, Y.; Wang, X.; Lin, L. Biomedical named entities recognition using conditional random fields model. In Proceedings of the International Conference on Fuzzy Systems and Knowledge Discovery, Xi'an, China, 24–28 September 2006; Springer: Berlin/Heidelberg, Germany, 2006; pp. 1279–1288.
18. Wang, D.; Li, Y.G.; Zhang, X.; Pu, X.Z. Chinese named entity recognition based on quasi-recurrent neural network. *Comput. Eng. Des.* **2020**, *41*, 2038–2043.
19. Li, Z.; Liu, F.; Yang, W.; Peng, S. A survey of convolutional neural networks: Analysis, applications, and prospects. *IEEE Trans. Neural Netw. Learn. Syst.* **2021**, *33*, 6999–7019. [CrossRef] [PubMed]
20. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, L.; Polosukhin, I. Attention is all you need. *Adv. Neural Inf. Process. Syst.* **2017**, *30*, 1–11.
21. Hochreiter, S.; Schmidhuber, J. Long short-term memory. *Neural Comput.* **1997**, *9*, 1735–1780. [CrossRef] [PubMed]
22. Dong, C.; Zhang, J.; Zong, C.; Hattori, M.; Hui, D. Character-based LSTM-CRF with radical-level features for Chinese named entity recognition. In *Natural Language Understanding and Intelligent Applications*; Springer: Berlin/Heidelberg, Germany, 2016; pp. 239–250.
23. Kim, Y. Convolutional neural networks for sentence classification. In Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing, Doha, Qatar, 25–29 October 2014; ACL: Stroudsburg, PA, USA, 2014; pp. 1746–1751.
24. Ma, X.; Hovy, E. End-to-end sequence labeling via Bi-directional LSTM-CNNs-CRF. In Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics, Berlin, Germany, 7–12 August 2016; ACL: Stroudsburg, PA, USA, 2016; pp. 1064–1074.
25. Devlin, J.; Chang, M.W.; Lee, K.; Toutanova, K. BERT: Pre-training of deep bidirectional Transformers for language understanding. In Proceedings of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Minneapolis, MN, USA, 16–21 June 2024; ACL: Stroudsburg, PA, USA, 2019; pp. 4171–4186.
26. Yu, Y.; Wang, Y.; Mu, J.; Li, W. Chinese mineral named entity recognition based on BERT model. *Expert Syst. Appl.* **2022**, *206*, 117727. [CrossRef]
27. Tong, Z.; Wang, L.D.; Zhu, X.J.; Du, Y. Research on named entity recognition in military domain based on pre-training model. *Front. Data Comput. Dev.* **2022**, *4*, 120–128.
28. Song, Y.; Wang, T.; Cai, P.; Mondal, S.K.; Sahoo, J.P. A comprehensive survey of few-shot learning: Evolution, applications, challenges, and opportunities. *ACM Comput. Surv.* **2023**, *55* (Suppl. S13), 1–40. [CrossRef]
29. Pourpanah, F.; Abdar, M.; Luo, Y.; Zhou, X.; Wang, R.; Lim, C.P.; Wang, X.Z.; Wu, Q.M.J. A review of generalized zero-shot learning methods. *IEEE Trans. Pattern Anal. Mach. Intell.* **2022**, *45*, 4051–4070. [CrossRef] [PubMed]
30. Cui, L.; Wu, Y.; Liu, J.; Yang, S.; Zhang, Y. Template-based named entity recognition using BART. In *Findings of the Association for Computational Linguistics*; ACL: Stroudsburg, PA, USA, 2021; pp. 1835–1845.

31. Shen, J.Y.; Tan, Z.Q.; Wu, S.H.; Zhang, W.; Zhang, R.; Xi, Y.; Lu, W.; Zhuang, Y. PromptNER: Prompt Locating and Typing for Named Entity Recognition. In Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics, Toronto, ON, Canada, 9–14 July 2023; Volume 1, pp. 12492–12507.
32. Ashok, D.; Lipton, Z.C. Promptner: Prompting for named entity recognition. *arXiv* **2023**, arXiv:2305.15444.
33. Layegh, A.; Payberah, A.H.; Soylu, A.; Roman, D.; Matskin, M. ContrastNER: Contrastive-based prompt tuning for few-shot NER. In Proceedings of the 2023 IEEE 47th Annual Computers, Software, and Applications Conference (COMPSAC), Torino, Italy, 26–30 June 2023; IEEE: New York, NY, USA, 2023; pp. 241–249.
34. Shao, J.X.; Huang, Q.; Xiao, C.; Liu, J.; Luo, W.B.; Wang, M.W. Two-Stage Prompt Learning for Few-Shot Named Entity Recognition. In Proceedings of the 23rd Chinese National Conference on Computational Linguistics, Taiyuan, China, 25–28 July 2024; Chinese Information Processing Society of China: Beijing, China, 2024; Volume 1, pp. 394–405.
35. Sang, E.F.; De Meulder, F. Introduction to the CoNLL-2003 shared task: Language-independent named entity recognition. In Proceedings of the Seventh Conference on Natural Language Learning at HLT-NAACL, Edmonton, AB, Canada, 31 May–1 June 2003; pp. 142–147.
36. Liu, J.; Pasupat, P.; Cyphers, S.; Glass, J. ASGARD: A Portable Architecture for Multilingual Dialogue Systems. In Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, Vancouver, BC, Canada, 26–31 May 2013; IEEE: New York, NY, USA, 2013; pp. 8386–8390.
37. Yang, Y.; Katiyar, A. Simple and effective few-shot named entity recognition with structured nearest neighbor learning. In Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing, Online, 16–20 November 2020; pp. 6365–6375.

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.