2022 21st International Symposium on Distributed Computing and Applications for Business Engineering and Science (DCABES)

Diagnosis of power operation and maintenance records based on pre-training model and prompt learning

Jun Jia
Jiangsu Electric Power Test and Research Institute
Electric Power Research Institute of State Grid Jiangsu Electric Power Co., Ltd
Nanjing 211103, China
jiajuntec@163.com

Hui Fu
State Grid Jiangsu Electric Power Co., Ltd
Nanjing 210000, China

Ziyang Zhang
State Grid Jiangsu Electric Power Co., Ltd
Nanjing 210000, China

Jinggang Yang
Jiangsu electric power test and Research Institute
Electric Power Research Institute of State Grid Jiangsu Electric Power Co., Ltd
Nanjing 211103, China

***Abstract*—The operation and maintenance records of power equipment contain abundant historical operation state information of equipment. However, due to the characteristics of multi ambiguity, difficult to segment ambiguity and multi noise, this paper proposes a two-stage model for the text diagnosis of power equipment. First, the large-scale pre-training model is trained based on the massive text, and then the pre-training language model is fine-tuned by the prompt technology for equipment diagnosis. The proposed solution is assessed through experiments and the numerical results demonstrate that the proposed solution can achieve about 20% improvement over the traditional method.**

Keywords- pre-training model; Natural language processing; deep learning; machine learning;

I. INTRODUCTION

In the power systems, a large number of unstructured data are distributed in the whole life cycle of design, installation, operation, maintenance and decommissioning, mainly including text, audio and images. Taking maintenance and repair as an example, power grid enterprises have accumulated a large number of maintenance test records, patrol inspection and defect elimination records, fault and defect description reports and event sequence records. These logs and reports are mainly in the form of short Chinese text (hereinafter referred to as text) mixed with numbers and alphabetic symbols. They contain rich historical operation status information, maintenance effect information and reliability information of equipment, which are very helpful to objectively evaluate the development process of equipment health status. However, the above information has not been fully mined due to the characteristics of multi ambiguity, difficulty to segment fuzziness and multi-noises.

Natural language processing (NLP) is a technology for systematic analysis, understanding and information extraction of text data intelligently and efficiently By using NLP and its components (e.g., [1, 2]), it is possible to manage very large blocks of power text data or perform a large number of automation tasks. The related technologies of natural language processing include word segmentation, named entity recognition, part of speech tagging, dependency syntax analysis, word vector representation technology, semantic similarity calculation, text analysis, and so forth.

From the technical point of view, there are still many difficulties in text information mining: on the one hand, the text belongs to unstructured data, and there is no clear expression form, which needs to be converted into computable and easy-to-understand structured data; On the other hand, natural language itself has fuzziness and ambiguity, and there is no unified judgment standard (i.e. loss function) for the accuracy of the results. In addition, text knowledge itself contains complex human thinking content such as logic, emotion and speculation, which is difficult to understand and use accurately. In the process of long-term operation, overhaul and maintenance of power equipment, a large number of unstructured text data such as defect and fault reports, test inspection records and repair and defect elimination documents have been accumulated. These data contain rich fault problem information, fault causes and maintenance methods and other key features. Mining the semantic information and causal relationship among them is of great significance for guiding the state evaluation and operation and maintenance of equipment.

In literature, much research effort has been made in this research area. The pre-training language model uses a large number of unlabeled corpus for training, which can obtain a more comprehensive representation of words and sentences and can be applied to different downstream tasks through fine-tuning [3]. Prompt is considered a supplementary description of different NLP tasks, which transforms different NLP tasks into cloze, completion, etc., making them similar to the tasks in the model pre-training stage, to fully mine the knowledge learned by the pre-training language model, so that the same pre-training language model can be applied to different tasks [4]. By constructing cues, the pre-training language model performs well in small sample text classification and conditional text generation tasks [4, 5]. Prompt tuning is to optimize the representation of prompts in the embedded space, i.e. continuous prompts, by training when the pre-training language model is frozen, to make it more consistent with specific tasks. It can not only improve the

2473-3636/22/$31.00 ©2022 IEEE
DOI 10.1109/DCABES57229.2022.00029

performance of the model but also reduce the labor cost of searching and designing prompts. This method greatly reduces the number of parameters that need to be optimized during training and makes it possible to reuse the pre-training language model. In recent years, how to construct and optimize prompts effectively has gradually become a research hotspot. The authors in [6] simplified the prefix tuning method, using the vocabulary or classification tags related to the downstream tasks to initialize the prompts, and the pre-training language model updates the hidden layer representation of the prompts according to the input context. This method updates fewer parameters than prefix tuning. In addition, the template is added with learnable virtual tags to enhance the presentation ability of the prompt template.

In addition, the study in [7] proposed a knowledge-aware prompt optimization method (know prompt) for relation extraction, which integrates the structural constraints between triples into the prompt template, and uses learnable virtual words to build prompts with knowledge injection, reducing the template construction cost and improving the template's perception of tasks. Although both methods use prompt tuning, they can only complete the task of relationship classification.

To this end, this work carries out the analysis and diagnosis of power operation and maintenance records based on the pre-training model and prompt learning. The main technical contributions made in this work can be summarized as follows:

(1) The prompt technology is applied to the text information extraction and diagnosis process of power operation and maintenance, and a new paradigm of power text diagnosis is formed based on the pre-training language model.

(2) A two-stage model reuse relational entity extraction model rept is proposed. In the first stage, fine-tune the pre-training language model to classify the relations contained in sentences; in the second stage, the pre-training language model fine-tuned in the first stage is optimized and reused by using prompts to extract and diagnose the text information of power operation and maintenance.

(3) The experiment is carried out on the power PMS data set. Compared with the traditional power operation and maintenance text information algorithm, the effect is improved by about 20%.

The rest of the work is organized as follows: Section II firstly formulates the problem and presents the proposed solution. Section III carries out the simulation experiments and presents the numerical results. Finally, the conclusive remarks are given in Section IV.

II. MODELS AND PROPOSED SOLUTION

Compared with the traditional "pre training fine tuning", prompt learning transforms the downstream tasks into language model tasks. In the pre-training process, the pre-training language model learns the knowledge information in the large-scale corpus through the self-supervised learning task, which suggests that learning makes the downstream tasks closer to the tasks in the model pre-training, and can fully tap the potential of the model. Compared with "pre-training fine tuning", It does not need to retrain a classifier parameter, saves the number of training parameters, makes the model convergence efficiency high, and is very effective in small sample learning.

A. Bidirectional Encoder Representation from Transformers(BERT)

Before diagnosing the equipment, the model needs to be pre-trained. The text classification task of the pre-training language model usually segments the input sentence first and then adds "[CLS]" and "[SEP]" marks before and after it. The final output hidden layer vector at the "[CLS]" position is used to represent the whole sentence, and then the classifier is used for classification. This paper uses the same method for relation classification. The input of the model is the sequence Ci, which is composed of a series of marks in the sentence, and the output is the probability estimate $\hat{y}_{rc} = \left[\hat{y}_{rc}^{1}, \hat{y}_{rc}^{2} \cdots \hat{y}_{rc}^{|\mathcal{R}|} \right]$ value that the current input contains a certain relationship.

In the model used in this paper, the power corpus is first input through the standard model, and then the input is word embedded. After multi-layer transformer coding, the last hidden layer output of each word in the relationship classification stage is finally obtained.

B. Prompt Working Process

Prompt is the technique of making better use of the knowledge from the pre-trained model by adding additional texts to the input. The new mode integrated with prompt can be roughly summarized as "pre-train, And predict ", in this mode, the downstream task is readjusted to a form similar to the pre-training task. For example, in the text emotion classification task, for the input of" power transformer oil leakage ", you can add prompt" ___ equipment oil leakage "after it In this way, by selecting the appropriate prompt, we can control the predictive output of the model, so that a completely unsupervised PLM can be used to solve various downstream tasks, thus achieving better results in the text diagnosis process of power equipment.

Prompt's workflow includes the following four parts: (1) structure of prompt template; (2) construction of prompt answer space map (verifier); (3) the text is substituted into the template, and the pre-training language model is used for prediction; and finally (4) map the predicted results back to the label. The details of the process are further described and discussed as follows:

Step 1: prompt construction (Template)

First, we need to build a template. The template is used to reconstruct the input and output into a new text with mask slots, as follows: define a template, including 2 slots filled in by generation: [x] and [Z]; and substitute [x] with the input text.

Step 2: answer construction (Verbalizer)

For the prompt constructed by us, we need to know the relationship between our prediction word and our label, and we can't run Z as an arbitrary word. Here we need a mapping function to map the output word with the label.

Step 3: answer prediction (Prediction)

At this point, we only need to select an appropriate pre-training language model and then predict mask slots [Z]. For example, in the following figure, we get the resulting fantasy, and we need to substitute it into [Z].

Step 4: answer-label mapping (Mapping)

For the obtained answer, we need to use the verifier to map it back to the original label.

C. Key Entity Extraction in Power Text

Cut the original text x = [x1, X2,..., xn] according to the n-gram method, construct the candidate keyword set K = [K1, K2,..., kn], respectively splice each element in the keyword set with the original text and the prompt template t = [T1, T2,..., TN] (where one element in the prompt template is [mask]), enter the Bert model as an input, predict the probability of [mask] content in the prompt template, and finally determine whether the element is a keyword.

$$L = -\sum_{i=1}^{n} logp(t_{mask} \mid X, k_i, T) \bullet label_{k_i} \qquad (1)$$

Where n is the number of candidate keywords.

The model diagram is as follows:：

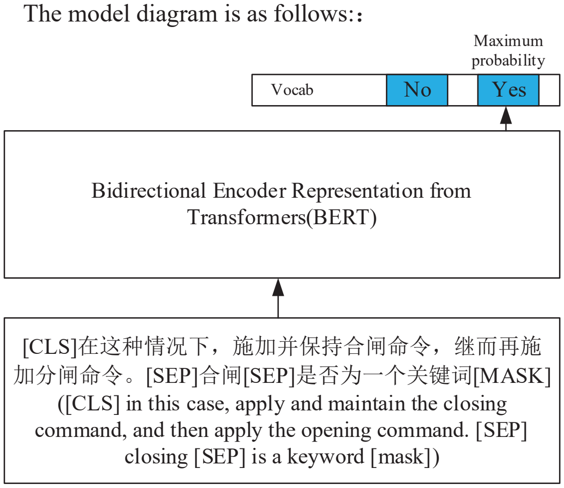

Fig 1. Prompt model diagram for two categories

As the additional information of the input, the prompt template can be divided into prefix, suffix or middle of the original input in terms of position, as follows:

TABLE I. PROMPT TEMPLATE

prefix	[CLS] 是否为一个关键词[MASK][SEP]在这种情况下，施加并保持合闸命令，继而再施加分闸命令。[SEP]合闸[SEP] (Whether [CLS] is a keyword [mask] [SEP] in this case, apply and maintain the closing command, and then apply the opening command. [SEP] closing [SEP])
Middle	[CLS]在这种情况下，施加并保持合闸命令，继而再施加分闸命令。[SEP] 是否为一个关键词[MASK] [SEP]合闸[SEP] ([CLS] in this case, apply and maintain the closing command, and then apply the opening command. [SEP] is a keyword [mask] [SEP] closing [SEP])
suffix	[CLS]在这种情况下，施加并保持合闸命令，继而再施加分闸命令。[SEP]合闸[SEP]是否为一个关键词[MASK] [SEP] ([CLS] in this case, apply and maintain the closing command, and then apply the opening command. [SEP] closing [SEP] is a keyword [mask] [SEP])

The output of the pre-training language model to the {mask} position in the template is the answer space, and the prediction result of the downstream task is the label space. The answer space mapping maps the answers in the answer space to the results required by the downstream task. We use simple answer space mapping "yes", "right", "wrong" and "no". "Yes" and "yes" indicate that the segment is a keyword; "False" and "no" indicate that the segment is not a keyword.

D. Fault Diagnosis based on Maintenance Records

The original text x = [x1, X2..., xn] is spliced with the prompt template t = [T1, T2..., Tn] (where one element in the prompt template is [mask]). It is input into the Bert model to predict the probability of [mask] content in the prompt template. Through the answer space mapping, it is judged that the text is general, serious and critical. The loss function is:

$$L = -logp\left(t_{mask} \mid X, T\right) \bullet label \qquad (2)$$

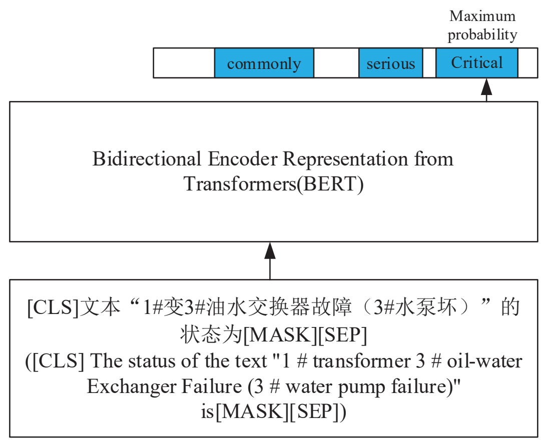

Fig 2. Prompt model diagram for three categories

E. Bidirectional Encoder Representation from Transformers(BERT)

Before diagnosing the equipment, the model needs to be pre-trained. The text classification task of the pre-training language model usually segments the input sentence first and then adds "[CLS]" and "[SEP]" marks before and after it. The final output hidden layer vector at the "[CLS]" position is used to represent the whole sentence, and then the classifier is used for classification. This paper uses the same method for relation classification. The input of the model is the sequence Ci, which is composed of a series of marks in the sentence, and the output is the probability estimate $\hat{y}_{rc} = \left[\hat{y}_{rc}^{1}, \hat{y}_{rc}^{2} \cdots \hat{y}_{rc}^{|\mathcal{R}|}\right]$ value that the current input contains a certain relationship.

In the model used in this paper, the power corpus is first input through the standard model, and then the input is word embedded. After multi-layer transformer coding, the last hidden layer output of each word in the relationship classification stage is finally obtained.

III. EXPERIMENTS AND NUMERICAL RESULTS

The fault diagnosis of power equipment is a typical text classification task. Based on the defect records in the maintenance records, combined with the information extraction and the relevant standards, determine the defect degree implied in the defect text. The examples of power defect diagnosis are shown in Table 1.

TABLE II. DEFECT DIAGNOSIS

Defect content	Defect classification
综合监控模块损坏(Integrated monitoring module is damaged)	commonly
2#主变本体瓦斯继电器处渗油(Oil leakage at gas relay of 2 # main transformer body)	commonly
3＃变本体呼吸器硅胶上层变色、整体硅胶潮解 1/2。(3 # transformer body respirator: the upper layer of silica gel is discolored and the whole silica gel is deliquescent (1 / 2))	serious
3#主变风冷控制开关故障合不上。(3 # main transformer air cooling control switch fails to close)	serious
1#变有载调压呼吸器油杯底部有冰，呼吸器不能呼吸。(There is ice at the bottom of the oil cup of 1 # transformer on load pressure regulating respirator, and the respirator cannot breathe.)	critical
10kVB 相套管发热 145 度。(10kvb phase bushing heating 145 degrees)	critical

A. Training set preparation

First of all, the data in the power defect data set is screened for the records with complete diagnosis information. After manual verification, 50%, 25% and 25% of the data are randomly cut. 50% is the training set, 25% is the verification set and 25% is the test set.

B. Experimental results

Compared with Bert, the improvement of prompt in small samples is higher than that in large samples. That is, with the increase of data volume, the advantage of prompt will decrease. However, compared with the original BERT, it is still improved, but the improvement is not obvious enough. Power text keyword extraction increased by 10% on 10 samples and 3% on 200 samples.

TABLE III. RESULTS OF PROMPT OF DIFFERENT SAMPLES

	BERT		Prompt	
F1 / Number of samples	Validation set	Test set	Validation set	Test set
10	0.3866	0.3749	0.4956	0.4982
100	0.4668	0.4856	0.5291	0.5309
200	0.5456	0.5597	0.5830	0.5894

The effect of the prompt is affected by the template. The effect before different templates is different. For the power text keyword extraction task, the suffix template has the best effect, while the template in the sentence has the worst effect.

TABLE IV. RESULTS OF PROMPT OF DIFFERENT TEMPLATE POSITION

	Prompt	
F1 / Template	Validation set	Test set
prefix	0.5756	0.5782
Middle	0.5691	0.5709
suffix	0.5830	0.5894

Power text fault diagnosis increased by 6% on 120 samples and 1% on 1200 samples.

TABLE V. RESULTS OF BERT WITH OR WITHOUT PROMPT

	BERT		Prompt	
Accuracy / Number of samples	Validation set	Test set	Validation set	Test set
120	0.4067	0.4307	0.4780	0.4988
720	0.5077	0.5264	0.5350	0.5488
1200	0.5657	0.5766	0.5796	0.6050

IV. CONCLUSIVE REMARKS

This paper proposes an analysis and diagnosis of power operation and maintenance records based on the pre-training model and prompt learning. The model adopts the architecture of relation classification first and entity extraction later, and realizes the reuse of two-stage pre-training language models by prompt. The experimental results show that the proposed solution is assessed and validated through experiments and the numerical results confirm the effectiveness of the proposed solution.

For future work, the proposed model needs to be further validated for a range of tasks and operational scenarios. In addition, the model can be further improved based on more advanced NLP models and algorithmic solutions.

ACKNOWLEDGMENTS

This work is supported by the Jiangsu electric power test and Research Institute Research on comprehensive performance evaluation of transformer based on multi-dimensional condition monitoring

REFERENCES

[1] Bordes, A., Glorot, X., Weston, J. et al. A semantic matching energy function for learning with multi-relational data. Mach Learn 94, 233–259 (2014).

[2] Everingham, M., et al., The Pascal Visual Object Classes (VOC) Challenge. International Journal of Computer Vision, 2010. 88(2): p. 303-338.

[3] Czapla, P., J. Howard and M. Kardas, Universal Language Model Fine-Tuning with Subword Tokenization for Polish. 2018.

[4] Brown, T.B., et al., Language Models are Few-Shot Learners. 2020.

[5] S., X. and F. W., Optimizing Continuous Prompts for Visual Relationship Detection by Affix-Tuning. IEEE Access, 2022. 10: p. 70104-70112.

[6] Lester, B., R. Al-Rfou and N. Constant. The Power of Scale for Parameter-Efficient Prompt Tuning. 2021.

[7] Chen, X., et al., KnowPrompt: Knowledge-aware Prompt-tuning with Synergistic Optimization for Relation Extraction. 2021.