A method of automatic generation of daily work report for power dispatching based on RoBERTa

Jie Li¹, Tao ZHANG², Hao LI², Jixiang LU²,³, Jinjun LU²,³, Ziqi LI²

¹ State Grid Jiangsu Electric Power Co., LTD
² Technology Research Center of NARI Group Corporation, Nanjing, China
³ State Key Laboratory of Intelligent Power Grid Protection and Operation Control, Nanjing, China

Lijie_tzgd@sina.com,
{zhangtao11,lihao5,lujinjun,lujixiang,liziqi}@sgepri.sgcc.com.cn

Abstract. Power dispatching operators need to read a large number of power dispatching logs every day and arrange them into daily work reports. It takes them a lot of time to do information extraction and transfer that information into work reports. In order to reduce the workload of power dispatching operators, we propose an intelligent method for automatic generation of daily work reports in this paper by adopting RoBERTa, which is an improved version of BERT. With the outstanding performance of RoBERTa in semantic analysis and entity recognition, the workload of power dispatching operators is reduced approximately 50% through our proposed method, which means the efficiency is greatly improved.

Keywords. Natural Language Processing, Information Extraction, Named Entity Recognition, Pattern Matching, RoBERTa Introduction

1. Introduction
Reading a large number of complicated logs generated from the power dispatching system and transferring them into daily work reports according to corresponding regulations is an essential part of the daily work for power dispatching operators. According to our preliminary investigation, it takes an experienced operator approximately two hours to read power dispatching logs and write daily work reports. In order to help power dispatching operators save time on log reading and information transferring, we propose an intelligent method for automatic generation of daily work reports based on corresponding regulations in this paper by utilizing some technologies and models of NLP based on deep learning [1, 2]. Nevertheless, similar to other NLP tasks, the biggest challenge in our task is the shortage of massive training data, since there are a lot of duplicate contents in power dispatching logs. It is hard to augment text data of logs as dispatching operation contents are strongly professional. Therefore, we choose transfer learning to solve this problem through benefiting from training on enormous amount of un-annotated text corpora, getting pre-trained general-purposed language model and fine-tuning the pre-trained model with specific sparse text data [3].
To make our proposed method more accurate, we choose RoBERTa [4] for semantic analyzing by using pre-trained general-purposed language encoders and fine-tuning them in power dispatching logs.
The experimental results show that the workload is reduced more than 50% by using our proposed method.

Our contributions can be summarized as follows:
1. To the best of our knowledge, what we present in this paper is the first work to exploit power dispatching logs contextualized embedding with RoBERTa pre-trained model, and improve word matching performance with fine-tuning.
2. We redesign a new model for named entity recognition (NER) by employing RoBERTa. It creates a state-of-the-art accuracy for named entity recognition task with the model RoBERTa-BiLSTM-CRF.

2. Related work

A lot of work has been done to make computers reach the level of human beings in terms of understanding human language. Word2vec is introduced and popularized by Mikolov et al. [5] on word embedding process. However, Word2vec cannot dynamically alter the word representation based on the contextual information. BERT, published by Google in late 2018, is a new method of pre-training language representations which obtains state-of-the-art results on a wide array of NLP tasks. Guo et al. [6] worked on reading answer from Database without SQL form, they propose method to learn from question and search result directly in tables with BERT based model. Zhang et al. [7] used BERT model on text summarization task successfully. Recently, the authors of BERT have released an updated version of BERT, which is called Whole Word Masking. The whole word masking mainly mitigates the drawbacks of original BERT that all the WordPiece tokens are masked altogether if a masked WordPiece token belongs to a whole word. It explicitly forces the model to recover the whole word in Masked Language Model (MLM) pre-training task instead of recovering WordPiece tokens [8]. RoBERTa, an improved recipe for training BERT models, is capable of matching or exceeding the performance of all of the post-BERT methods [9]. RoBERTa is the model that we adopted in our proposed method as it proves upon the published BERT results on both GLUE and SQuAD in terms of controlling for training data.

3. Implementation

3.1. Overview of the method

In this paper, we propose an intelligent method of handling power dispatching logs based on NLP technologies. The proposed method is able to assist operators to arrange daily work reports and improve working efficiency. As shown in Figure 1, the whole process is composed of three steps.

![Figure 1](image)

**Figure 1.** The whole process of the intelligent method.

**Step 1:** We pre-process the original logs to organize them into records with timestamps. Key contents that we extract from the logs are “Dispatching Operation”, “Electrical Overhaul”, “Power Grid Accident”, “Electrical Defect”, “Start-up and Shutdown” and “Operation Mode Adjustment”.

**Step 2:** We analyze the regulations, extract mode elements such as objects, conditions and events from the regulations and form regular patterns composed of mode elements. For example, if we need to
parse the sequence “Start-up of devices with the voltage level of 220kV and above in a certain province”, we would come up with a pattern “Status + Device + Voltage Level + Province”. In order to enhance accuracy, detailed regulation is adopted to define types for those words extracted from daily work reports. The operation function is constructed based on the corresponding type. **Step 3:** We utilize NLP technologies to do word segmentation on power dispatching logs based on domain dictionaries. Dictionary based word type annotation is carried out at the same time during word segmentation. Besides, NER is used to get the pattern element annotation of each word. With the annotations of word type and pattern element, sentences are matched into different patterns. In particular, when dealing with those logs lack of conditions, we make supplement by querying domain knowledge. Finally, we extract contents that meet the rules and classify them into different categories through objects. Contents with the same object are merged into one record.

3.2. Key Technologies

3.2.1. Chinese Word Segmentation

We use Jieba to cut specific text sequences into segmented words. The accuracy of word segmentation is highly dependent on the dictionary for Chinese. Therefore, we have done a lot of work building a comprehensive and reliable dictionary. During the work of dictionary building, we have summarized that major device types of power dispatching include substation, power station, line, etc. The names of specific devices are provided by dispatching departments of electric power companies. Also, we have done some research on abbreviations and aliases of those standard device names.

3.2.2. Word Based Name Entity Recognition

Name Entity Recognition (NER) is one of the most essential NLP tasks, aiming at identifying entities and laying the groundwork for entity relation extraction. After researching on BERT, we come up with the conclusion that the model pre-trained with Whole Word Masking enjoys better performance than that pre-trained with WordPiece Masking for Chinese. As a result, we adopt word based NER models. The classical NER model is Word2vec-BiLSTM-CRF. As what is shown in Figure 2(a), the model architecture of Word2vec-BiLSTM-CRF contains three layers.

![Figure 2. The model architecture of Word2vec-BiLSTM-CRF and RoBERTa-BiLSTM-CRF.](image)

The first layer is a looking-up layer, in which each word $x_i$ of a sentence is transferred from a one-hot vector into a low density word vector through pre-training or randomly initialized embedding matrices. The second layer is a BiLSTM layer in which sentence features are automatically extracted. After the feature extraction, the hidden layer state vectors are mapped from m dimensions to k dimensions through a linear layer after dropout where k is the number of labels of marked data sets. Thus, sentence features are successfully extracted as $P \in R^{m \times k}$, where $P_{ij}$ is regarded as the score for the
mapping from category of the word $x_i$ to the $j$-th label. The third layer is a CRF layer used for sentence-level sequence labelling. The parameter of the CRF layer is a $(k+2)^*(k+2)$ matrix $A$. A specific member $A_{ij}$ is the score for the transferring from the $i$-th label into the $j$-th label. Therefore, the previous label can be used when marking a position.

In terms of Word2vec, the biggest drawback is its poor effect on ambiguous words. For example, the word “apple” is only represented as one vector in Word2vec despite of the context. Thus, it may be figured out as a kind of fruit or a name of a hi-tech company. Compared with Word2vec, RoBERTa, the abbreviation of Robustly Optimized BERT Pre-training Approach, is an upgraded version of BERT and capable of representing different vectors for different meanings. Thus we propose a new model with the combination of BiLSTM, CRF and RoBERTa for NER tasks. The architecture is shown in Figure 2(b).

4. Experiment

4.1. Data Preparation

Limited by data confidentiality, we use approximately 4000 sentences from power dispatching logs of two weeks provided by the dispatching department of a provincial electric power company. Those sentences are segmented words by words and the type of each word is marked. As shown in Figure 3, in terms of marking for word type, “Substation” and “Line” are jointly marked as “S_S”, “Switch” is marked as “S_O”, and “Status” is marked as “S_A”. In order to prevent over fitting, data augmentation is adopted. We implement data augmentation by replacing words, which means that an entity is randomly replaced by another entity of the same area with the same voltage level and the same type. Hence, the data is expanded 20 times.

![Figure 3. Annotated case for pattern matching.](image)

4.2. Word Segmentation

We make use of Jieba to do word segmentation on the expanded data set, and compare the precision of word segmentation based on dictionaries with that without any dictionary. The precision is calculated as the ratio of the number of correct words segmented by the machine to the total number of manually segmented words. The result is shown in Table 1.

| The way of word segmentation | Precision (%) |
|-----------------------------|--------------|
| With dictionaries           | 91.3         |
| Without dictionaries        | 70.8         |

4.3. NER

For the experiment of NER, we divide the expanded data into the training data and the test data in a ratio of 8:2 after data augmentation is done. We make use of three models for NER training in three experiments. The results are shown in Table 2. It can be seen from the experimental results that RoBERTa-BiLSTM-CRF enjoys better performances than Word2vector-BiLSTM-CRF. Besides, RoBERTa-BiLSTM-CRF with fine-tuning is better than that without fine-tuning. We come up with two viewpoints after analyzing the results. First of all, Word2Vec does not do well in distinguishing named entities of the same word in different sentences, which makes it easy to make mistakes when marking the words. Moreover, RoBERTa, a pre-trained model for Chinese, is trained based on common language. Therefore, a fine-tuned one is better for the corresponding specific domain.
Table 2. NER training and testing results.

| Category | Index   | Word2vec (%) | Roberta without fine-tuning (%) | Roberta with fine-tuning (%) |
|----------|---------|--------------|---------------------------------|------------------------------|
| training | Precision | 86.7         | 90.8                            | 95.3                         |
|          | Recall   | 81.9         | 85.6                            | 90.5                         |
|          | F-measure | 84.23        | 88.12                           | 92.84                        |
| testing  | Precision | 85.5         | 89.3                            | 94.5                         |
|          | Recall   | 80.2         | 83.9                            | 89.8                         |
|          | F-measure | 82.77        | 86.52                           | 92.09                        |

5. Conclusion and Future Work

The old way of generating daily work reports is to extract information from power dispatching logs through regular expressions. The biggest shortcoming of this way is that it takes a lot of manpower to summarize regular rules before regular matching. In order to solve this problem, we propose an intelligent method to generate daily work reports of power dispatching based on NLP technologies in this paper. RoBERTa is adopted as the semantic analytic model in our proposed method, which reaches a state-of-the-art accuracy of particular domain NLP task. Our proposed method significantly improves efficiency by reducing approximately 50% of the workload and ensures that all the contents complied with rules will be extracted even though there is redundancy.

In terms of future work, we need to focus on some advanced technologies such as data correcting and knowledge reasoning to improve the accuracy of object element recognition and reduce redundant data. When accuracy is improved and data redundancy is solved, our proposed method shall be more universally adaptive.

Acknowledgments

This work is supported by National Key Development Plan Program (No.2016YFB0902600) and State Grids Science and Technology Program (Research on Key Technologies of Knowledge Graph in Power Grid Fault Disposal), and partially supported by Collaborative Innovation Center of Novel Software Technology and Industrialization. This work is supported by Science and Technology Program of NARI Technology Development Limited Company (Research and Application of Key Technologies in the Construction of Power Dispatching Knowledge Graph).

References

[1] Zhou Niancheng, et, al. Developing Status and Prospect Analysis of Deep Learning in Smart Grid. [J]. Automation of Electric Power Systems, 2019,43(04):180-197
[2] Shan Xin, et al. Analysis of Key Technologies for Artificial Intelligence Applied to Power Grid Dispatch and Control [J]. Automation of Electric Power Systems, 2019,43(01):49-57
[3] Cheng Lefeng, et al. Machine Learning for Energy and Electric Power Systems State of the Art and Prospects [J]. Automation of Electric Power Systems, 2019,43(01):15-43
[4] Devlin J, Chang M W, et al. Bert: Pre-training of deep bidirectional transformers for language understanding [J]. arXiv preprint arXiv:1810.04805 (2018).
[5] Mikolov T, Sutskever I, Chen K, et al. Distributed representations of words and phrases and their compositionality[C]//Advances in neural information processing systems. 2013: 3111-3119 (2013).
[6] Guo T, Gao H. Table2answer: Read the database and answer without SQL [J]. arXiv preprint arXiv:1902.04260 (2019).
[7] Zhang H, Gong Y, Yan Y, et al. Pretraining-Based Natural Language Generation for Text Summarization [J]. arXiv preprint arXiv:1902.09243 (2019).
[8] Cui, Yiming , et al. "Pre-Training with Whole Word Masking for Chinese BERT." (2019).
[9] Liu, Yinhan , et al. "RoBERTa: A Robustly Optimized BERT Pretraining Approach." (2019).