Research on power grid scheduling log word vector extraction based on bidirectional LSTM combined dictionary

Xiaohui Pan¹, Tan Chen² and Shengnan Liu²*

¹State Grid Jiangsu Electric Power Co., Ltd., Nanjing, Jiangsu, 210000, China
²Beijing Kedong Electric Power Control System Co., Ltd., Beijing, 100192, China
*liushengnan1@sgepri.sgcc.com.cn
*Corresponding author’s e-mail: 524869920@qq.com

Abstract. With the rapid development of China’s economy, the power network specifications are expanding and the network structure is becoming more and more complex. Power grid dispatching is the key to ensure the safe and stable operation of power grid. Power grid dispatch log is an important data source to reflect the operation of power grid and an important means to monitor the daily operation of power grid. Network dispatching log classification is an important application of log text analysis and mining. At present, there are many methods for network dispatching log classification, including naive bayesian method, support vector machine, neural network model and so on. However, no matter what classification method is used, scheduling log text needs to be preprocessed and converted into vector form before model training and classification. At present, the research of word vector mainly focuses on the Internet, while the feature extraction of power grid dispatch log from word vector generation is less. In this paper, a method of extracting log word vectors from power grid dispatching based on bidirectional LSTM combined dictionary is proposed. Firstly, the original log is preprocessed according to the lexicon, and word segmentation is performed on the original log by means of bidirectional LSTM combined with dictionary to obtain word segmentation results. Then, every word is transformed into a word vector through the skip-gram model. Finally, the generated word vector is used to classify the power grid dispatch logs.

1. Introduction
The level of economic development in our country has gradually improved, and the demand for electricity is also increasing. The installed capacity of power systems in corresponding regions has also increased year by year. The scale of the power grid has expanded year by year, and the network structure has become increasingly complex. The key to ensuring the stable operation of the power grid system is the power grid dispatching, and how to better monitor the operation of the power grid dispatching, and thus improving the management level of power dispatching, is a problem that power companies urgently need to solve.

At this stage of the power grid, there are many types and huge numbers of power equipment. When one of the power equipment fails, the power grid dispatcher will record the name of the failed plant, the interval, type of equipment, name of equipment, voltage level of equipment, failure phenomenon of equipment and failure reason of equipment. However, due to the different log recording habits of each dispatcher and the large difference in professional level, the recording methods of grid dispatch logs are also very different. When the power grid system is running, corresponding scheduling
information will be generated and stored in the form of logs. By analyzing and mining these dispatch logs, grid dispatchers can obtain a lot of valuable information that reflects the operating status of the grid. Grid dispatch log classification is an important application for log text analysis and mining. Current grid dispatch log classification methods are numerous, including naive Bayes [1], support vector machine [2], neural network [3] and so on. But regardless of the classification method, the scheduling log needs to be processed first and converted into a vector form.

When dealing with natural language, the computer can only convert it into a binary coded form, called a vector. The conversion of word vectors can usually be achieved by two methods, one is the one-hot vector [4] method, and the other is the word embedding [5] method. One-hot vector expression is simple and has a significant effect on traditional text classification. The disadvantage is that the generated word vector has a large dimension and cannot reflect the relationship between related words. In order to solve this problem, the word vector came into being. It not only solved the dimensional disaster problem of the one-hot vector, but also retained rich contextual semantic information. In tasks such as text classification, text retrieval, machine translation, and many other natural language processing fields, word vectors are widely used for text feature representation, and in most tasks, better performance than single hot vectors is obtained.

At this stage, in the field of log analysis, the effect of word vectors is more significant. Li Youxu et al. [6] constructed a neural network model and used word vectors as input to diagnose faults by analyzing network logs. Zhang Chi et al. [7] proposed a new text clustering method based on word vectors. Xu Jiahui [8] and others used word vector training to generate a GRU neural network model to classify the alarm information of the power grid. At present, the research is mainly focused on the Internet field, and the feature extraction for the power grid dispatch log is less work in the generation of word vectors.

The content recorded in the grid dispatch log is diverse, rich in information, and has a wide range of sources. Based on the above characteristics, this paper proposes a method for extracting the word vectors of the power grid dispatch based on the bidirectional LSTM model and combined with the power system professional dictionary. First, perform word segmentation on the original log through the bidirectional LSTM model, and match through the professional lexicon of the power system to obtain the word segmentation results. Afterwards, convert each word into a word vector through the skip-gram model. Finally, use the generated word vector, respectively input Naïve Bayes (NB) model, K-neighborhood (KNN) model, support vector machine (SVM) model, period memory network (Long Short-Term Memory, LSTM) model for training, and then use each model to classify the scheduling logs and view the results. Experimental results show that this method has better effect than traditional word segmentation method.

2. Grid dispatch log

2.1. Overview of grid dispatch log
The content of the power grid dispatch log includes information such as whether the power grid is currently operating, whether there is a fault, the fault phenomenon, the cause and the resolution process, the implementation of the grid's routine maintenance plan, the emergency events in the power grid, and the head office and branches. These logs are recorded in the form of natural language, and the length is uncertain, which contains various information such as symbols, time, names and places.

Taking the provincial power company's 2017 grid dispatch log as an example, it contains 7,055 valid records. Excluding the date and time in each record, the shortest record contains 20 characters and the longest record contains 503 characters. The lengths vary greatly. After eliminating invalid samples, manually labeling log categories, word segmentation and other data preprocessing steps, a total of 4,603 data categories were formed, including wind power generation, photovoltaic power generation, hydropower generation, and power grid operation.

For example, a grid dispatch log labeled "wind power", the original log is as follows: 2017-10-01 14:57:02 Pfeiffer Central Control (Fu Xiaojian) reported: 14:40 The fuse in the C-phase box
transformer of No.26 fan in Jufeng wind farm broke, apply for shutting down all 11 fan stations #23-33 along No.3 fan line controlled by 313 switch, the total capacity is 16.5MW. Dispatch in north Hebei agreed. 15:02 the above fan shutting down over.

Pre-processing results: "Jufeng, wind farm, fan, box transformer, fuse, broke, switch, fan line, fan station, shutting down, total capacity, north Hebei, dispatch, unified, above, fan shutting down, over”.

2.2. Characteristics of grid dispatch log

The grid dispatch log is a text that records the operation of the grid system. It has the following characteristics:

1) There is no fixed format for the log text, and there are various forms. Grid dispatching log records are recorded by dispatching personnel at all levels of power companies. Due to their different cultural levels and recording habits, the form of log records is also different, which has certain interference with the effect of word segmentation.

2) The grid dispatch log contains a large amount of professional vocabulary for grid dispatch, which also contains a lot of time, symbols, formulas, units, etc.

The dictionary-based word segmentation method has strong domain adaptability, but the word segmentation efficiency is poor, and dictionary maintenance is difficult. Machine learning algorithms based on statistics are less effective in professional applications, but have high efficiency. Therefore, the combination of statistical-based machine learning algorithms and dictionary-based word segmentation method can be very good for word segmentation processing of professional field text.

3. Bidirectional LSTM combined dictionary word segmentation method

In terms of Chinese word segmentation, there are currently three main methods, which are word segmentation based on string matching, word segmentation based on understanding and word segmentation based on statistics. This article uses a bidirectional LSTM combined dictionary word segmentation method to process the scheduling log, and its work flow chart is shown in Figure 1:

![Figure 1. Flow chart of bidirectional LSTM combined dictionary word segmentation method.](image-url)
The grid dispatch log is very suitable for word segmentation based on understanding because of the particularity of its field. In this paper, for the transfer of logs, a knowledge base of special words for grid dispatch is first defined, and then a two-way LSTM neural network model is built, and the model is used to segment the dispatch log. Regarding the output word segmentation results, when the bidirectional LSTM model cannot accurately segment words, the words are passed to the expert system for further judgment. This method can further improve the accuracy of word segmentation.

LSTM (Long Short-Term Memory) is a kind of neural network model, but it is special. Unlike the traditional RNN model, it can avoid the gradient disappearing. The structure of the LSTM model is shown in Figure 2:

![Figure 2. LSTM model structure.](image)

In the LSTM model, at each sequence index position \( t \), there will generally be three gates, specifically an input gate, an output gate, and a forget gate. The input gate is used to process the input information of the current sequence position; the output gate is used to process the output information of the current sequence position; the forget gate is used to control whether the current sequence position is forgotten, and the LSTM model will control whether to forget the upper hidden cells with a certain probability status. Bidirectional LSTM (BiLSTM) combines the forward LSTM model and the backward LSTM model. Its structure is as follows:

![Figure 3. Bidirectional LSTM model structure.](image)

As can be seen from Figure 3, the output layer is connected above and below the Forward layer and the Backward layer. The model structure contains a total of six shared weights, namely \( W_1, W_2, W_3, W_4, W_5, W_6 \). The model construction includes two processes. First, in the Forward layer, the forward calculation is performed from time 1 to time \( t \), and the output of the forward hidden layer at each time is saved. Afterwards, in the Backward layer, reverse calculation is performed from time \( t \) to time 1, and the output of the backward hidden layer at each time is saved. Finally, at each moment, the output of the Forward layer and the output of the Backward layer are combined to obtain the final output. The formula is as follows:

\[
h_t = f(w_1 x_t + w_2 h_{t-1})
\]
\[ h'_t = f(w_3x_t + w_5h'_{t+1}) \]  
\[ o_t = g(w_4x_t + w_6h'_t) \]

4. Improved skip-gram model to generate word vector

4.1. Introduction to skip-gram model

Word2vec is a shallow neural network model that can be used to generate word vectors. There are currently two implementation models for this method. The first is a skip-gram model. This method uses the head word to infer the words in a certain window of the context. The second one is the continuous bag of words (CBOW) model. Contrary to the skip word model, this method is to infer the central word through the words in a certain window below the line [9].

The skip-gram model is a commonly used word vector generation model. It mainly includes two parts, which are the training model and obtaining embedded word vectors. In the training model stage, a neural network model is constructed from the training data. When the model training is completed, the model will not be used to work. In the stage of obtaining the embedded word vector, the weight matrix of the hidden layer of the trained neural network is obtained. It is the last required word vector [10].

The skip-gram model structure is shown in Figure 4.

![Figure 4. skip-gram structure model.](image)

In the skip-gram model training process, the training samples need to be converted into the form of one-hot vector, so before training the model, we need to get the vocabulary of the training data [11]. The conditional probability is obtained by performing the inner product operation softmax on the vector:

\[ P(\omega_o | \omega_c) = \frac{\exp (\mu_o^T u_c)}{\sum_{i \in V} \exp (\mu_i^T u_c)} \]  

Where, each word is represented by two d-dimensional vectors to calculate the conditional probability.

In the skip-gram model, each word is represented as two d-dimensional vectors, which are used to calculate the conditional probability. \(i\) is the index of this word in the dictionary. When it is the head word, the vector is expressed as \(u_i \in \mathbb{R}^d\), and when it is the background word, the vector is expressed as \(\mu_i \in \mathbb{R}^d\).
4.2. Improving skip-gram model

The traditional skip-gram model has no requirements on the order of input words, that is, the order of input words has no effect on the generation of the final word vector. However, for the grid dispatch log, the order of words is really important, because the record of a grid dispatch log may contain multiple keywords, including both abnormal keywords and normal keywords. This is because a grid dispatch log will record the entire process, which is very important for the final classification judgment.

In order to solve this problem, this paper improves the traditional skip-gram model. For the traditional skip-gram model, we will first select the input word, and then define the window parameter, that is, skip window. The meaning it represents is the number of words selected from the side of the input word. Then, we finally get the number of words in the window up to 2 * skip window, the words in the window are called output words, then the training format we finally get is (input word, output word). Using these training samples as input to train the neural network model, it will eventually output a probability distribution, which represents the possibility that each word in the dictionary is an output word, and the hidden layer weight W is the word vector that is sought.

The difference between the improved skip-gram model and the traditional skip-gram model is that the output layer uses hierarchical softmax instead of softmax. Therefore, in the mapping process from the hidden layer to the output layer, this paper creates an output for each output vocabulary. The matrix is then mapped to each target word according to this output matrix. The structure of the improved skip-gram model is shown in Figure 5:

![Improved skip-gram model structure](image)

Figure 5. Improved skip-gram model structure.

Among them, w1, w2, etc. are the output matrix of each word. The function softmax will construct a Huffman tree according to the word frequency of the words in the dictionary, then calculate the probability problem of the current word in its context, and then convert it to the path prediction problem in the Huffman tree, that is, a binary classification problem, using logistic regression to calculate. Specifically:

The probability of being divided into the right subtree is:

\[ P(\text{right subtree}) = \frac{1}{1 + e^{-x_W^T \theta}} \]  

(5)

The probability of being divided into the left subtree is:

\[ P(\text{left subtree}) = 1 - P(\text{right subtree}) \]  

(6)

Among them, \( x_W^T \) is the word vector of the current internal node, and \( \theta \) is the model parameter of logistic regression that we need to compute from the training sample.

The traditional skip-gram model predicts the context based on the head word. Assuming that when the window size is d, if the head word is at the beginning or end of the sentence and there are not so many words in the window, then the size of d will adjust accordingly. In other words, when the window is defined as d, the actual window size is a random number from 1 to d. Because of the
random variation of the window size, the improved skip-gram model retains the mechanism of the original model, but adds 0 to 2d coefficients.

5. Experiment and analysis

5.1. Experiment design and evaluation index
In this paper, 4603 dispatch logs of a provincial power company are used as training data and test data, and cross-validation is used to train modules and tests. The ratio of training data to test data is 8:2. After the word segmentation operation is performed on the dispatch log, the bidirectional LSTM combined with the word segmentation method of the dictionary and the traditional word segmentation method are used to convert the transferred logs into word vectors, and then use them as training data to train the Naive Bayes classifier, SVM classifier, LSTM classifier, to compare the dispatch log classification effect when two different word vectors correspond to the same classifier.

The evaluation indicators for the classification effect of the classifier include precision, recall, and F1 value. The calculation method of each evaluation index is as follows:

| Actual situation | Forecast result | | |
|------------------|-----------------|-----------------|
| Positive example | TP (True positive example) | FN (False negative example) |
| Negative example | FP (False positive example) | TN (True negative example) |

The calculation method of precision rate is as follows:

\[
P = \frac{TP}{TP + FP}
\]  
(7)

The calculation method of recall rate is as follows:

\[
R = \frac{TP}{TP + FN}
\]  
(8)

The calculation method of F1 value is as follows:

\[
F1 = \frac{2 \times P \times R}{P + R}
\]  
(9)

5.2. Experiment results and analysis
The experiment results are shown as Table 2.

| Algorithm | Category label | Traditional word segmentation method | Bidirectional LSTM combined with dictionary word segmentation method |
|-----------|----------------|--------------------------------------|---------------------------------------------------------------|
|           |                | Precision rate | Recall rate | F1 value | Precision rate | Recall rate | F1 value |
| NB        | Wind power generation | 75.6% | 76.2% | 82.3% | 84.2% |
| NB        | Photovoltaic power generation | 69.5% | 65.4% | 77.2% | 78.9% | 75.8% |
The experimental results are shown in Table 2. As can be seen from Table 2, compared with the traditional word segmentation method, the word vector generated by the bidirectional LSTM combined with the dictionary word segmentation method can significantly improve the classification effect of the scheduling log classifier. Whether it is an NB classifier, SVM classifier, or LSTM classifier, compared to the word vector generated by the traditional word segmentation method, when the classifier uses the word vector generated by the bidirectional LSTM combined with the word segmentation method of the dictionary, the precision rate, recall rate and F1 value of different types of classifiers have been significantly improved, indicating that compared with the traditional word segmentation method, the proposed bidirectional LSTM combined dictionary-based grid dispatch log word vector extraction technology has a better classification effect.

6. Conclusion
Aiming at the characteristics of power grid dispatch logs, this paper proposes a bidirectional LSTM combined dictionary word segmentation system, and improves the traditional skip-gram model to obtain more reasonable word vectors to adapt to the characteristics of power grid dispatch logs. Through comparative experimental analysis, the proposed method for extracting word vectors of power grid dispatch logs based on bidirectional LSTM combined with dictionaries can effectively improve the rationality of word vectors, which has a great impact on the final power grid dispatch log classification and has a significant effect.

References
[1] Wang, Y.Y. (2019) Application of naive bayes method in chinese text classification. J. China High-Tech, 43(07):59-62(in Chinese).
[2] Zhang, M.Y., Ai, X.B., Hu, Y.Z. (2019) Chinese text classification system on regulatory information based on SVM. J. IOP Conference Series: Earth and Environmental Science, 252:022133.
[3] Yang, M., Zhao, W., Chen, L., QU, Q., Shen, Y. (2019) Investigating the transferring capability of capsule networks for text classification. J. Neural Networks, 118.
[4] Pham, D.H., Le, A.C. (2018) Exploiting multiple word embeddings and one-hot character vectors for aspect-based sentiment analysis. J. International Journal of Approximate Reasoning, S0888613X17304139.

[5] Nguyen, T.T.D., Le, N.Q.K., Ho, Q.T., Phan, D.V., Ou, Y.Y. (2019) Using word embedding technique to efficiently represent protein sequences for identifying substrate specificities of transporters. J. Analytical Biochemistry, 577.

[6] Li, Y.X. (2017) Network fault diagnosis model based on convolutional neural network. J. Software Guide, 16(12):40-43(in Chinese).

[7] Zhang, C., Zhang, G.H. (2019) Text clustering algorithm based on word vector and multi-feature semantic distance. J. Journal of Chongqing University of Science and Technology (Natural Sciences Edition), 21(03):69-72+77(in Chinese).

[8] Xu, J.H., Zhang, H., Xiao, L.P., He, H., Zhang, Y., Geng, Y., Zhou, Y.S. (2019) Research on classification of power grid alarm information based on GRU neural network. J. Computer and Digital Engineering, 47(06):1405-1408+1538(in Chinese).

[9] Zheng, W.C., Xu, P. (2013) Research on clustering Chinese words with word2vec. J. Computer and Digital Engineering, 34(12):160-162(in Chinese).

[10] Zhang, J., Qu, D., Li, Z. (2015) Recurrent neural network language model based on word vector features. J. Pattern Recognition and Artificial Intelligence, 28(04):299-305(in Chinese).

[11] Devi, G.R., Veena, P.V., Kumar, M.A., Soman, K.P. Fan, G.M., Guo, L.X. (2016) Entity extraction for malayalam social media text using structured skip-gram based embedding features from unlabeled data. J. Procedia Computer Science, 93:547-553.