Forecast Method of Energy Public Opinion Based on Conditional Random Field

Yu Wang¹, Shengzhen Xie², Junri Tang², Ling Sui², Kaiqi Kou² and Zhiqiong Wang¹,*

¹School of Medicine and Bioinformatics Engineering, Northeastern University, Shenyang, China
²School of Computer Science and Engineering, Northeastern University, Shenyang, China

*Corresponding author email: wangzq@bmie.neu.edu.cn

Abstract. As the driving force of development in the new century, the energy problem is an important factor restricting economic development, which is related to a country's economic security and national security. The public's views or attitudes towards (heat events related to energy issues, such as the State Grid's repeated public attention during the 2016 electricity reform) significantly affect the government's monitoring and guidance of the oil crisis. Ordinary research on emotion prediction regards the change of public attitude towards events as a random process, only considering the influence of the attitude tendency held by users in the early stage on the later stage of emotion development, and not considering the influence of external factors on emotion prediction in the actual situation. This paper proposes a public opinion emotion prediction method based on conditional random fields. Taking the transition probability between external factors as the characteristic transition matrix of the conditional random field, considering the influence of external factors on the development of events, and combining with bidirectional LSTM, the BILSTM-CRF model in this paper is proposed. Experiments show that the emotion prediction method based on conditional random fields can predict the emotion state more accurately, which verifies the superiority of the emotion prediction method based on conditional random fields.

1. Introduction
With the vigorous development of the Internet and the energy project, users can freely communicate on the Internet and express a large number of comments and opinions on energy-related issues on various social platforms on the Internet. The energy industry is basically in a monopoly state. Related enterprises and events are easily repugnant to public opinion. Therefore, it is extremely necessary to analyze the public opinion on the heat events of energy-related news. Emotional analysis, also known as opinion digging and opinion analysis. As a core work of emotion analysis, emotion classification is mainly aimed at automatic classification of emotions contained in texts. At present, there are two commonly used emotion classification methods, one is based on emotion dictionary and the other is based on machine learning. The method based on sub-points is mainly to manually specify a series of emotional dictionaries and rules, to split, analyze and match the dictionaries of the text, and finally to use emotional values as the basis for judging the emotional tendency of the text. The method based on machine learning regards emotion classification as a common classification problem, and the most important problem is feature extraction: commonly used text feature extraction is generally based on tf-idf and vector representation, followed by commonly used machine
learning algorithms such as support vector machine and Naive Bayes for emotion classification. These two methods require a large amount of manual annotation data and rely too much on background knowledge, and their ability to capture deep semantic information is still lacking. In summary, the traditional classification model of affective dictionaries needs a lot of prior knowledge as the background and has some deficiencies in semantic understanding. LSTM neural network can capture the context information more completely, and CRF (Conditional Random Field) layer can add constraints such as external influencing factors to ensure the final prediction results are more accurate. Therefore, based on word2vec, LSTM neural network and conditional random field, a new method model for public opinion emotional analysis of energy issues is proposed.

2. Related Work

2.1. Word2Vec Model

Word2vec model is a tool developed by Mikolov et al. On the basis of Log-Bilinear and NNLM models [7]. Using word2vec words can be mapped from high-dimensional space to low-dimensional space in a distributed way and the positional relationship between union quantifier vectors can be well preserved, thus solving the problems of vector sparsity and semantic connection. It can be divided into CBOW [8] (predicting the central word through the nearby word) and Skip-gram (predicting the nearby word through the central word). The training of CBOW model is to output the word vector of the word corresponding to the context of a certain word and predict the word vector of this characteristic word. Skip-gram model is the opposite of CBOW. In this paper, we use the CBOW model, which gives the context vector content (w) of the word w, and then predicts w according to the content (w). If the specified context (w) is assumed to be a positive sample, then the rest is a negative sample, and the negative sample set U (w) about w can be obtained by negative sampling [9].

2.2. LSTM Brief Introduction

LSTM [10] is derived from RNN. RNN can learn to use the previous information. However, when the relevant information is far away from the current predicted position, RNN will lose the ability to learn the information far away, and it is also difficult to learn. LSTM is specially designed to solve such a long-standing problem. LSTM is deliberately designed to avoid long-term dependency problems. The LSTM repeating module contains four interactive layers. The amnesia stage indicates which part of the data the neuron discards, the input stage indicates which data should be input, and the output stage indicates what information is output. The LSTM structure diagram is shown in Figure 1.

2.3. CRF Brief Introduction

CRF [11] is a discriminant model combining maximum entropy model and hidden Markov model (Hidden Markov Model) model. The learning and prediction of CRF model are generated according to the features of samples. Features can be selected according to the set feature template, and the optimal
solution can be obtained by optimizing all feature weights, thus solving the problem that HMM model cannot consider its context features due to the assumption of independence.

3. Emotion Prediction Method Based on Conditional Random Field

This paper proposes a depth model of emotion classification based on Word2vec, LSTM and CRF. Firstly, the existing expected news information is processed, and then the word vector of the news information is obtained by using Word2VEC tool, which is used as the input of LSTM network model designed by text. After linear transformation of features and Softmax function, the obtained probability is used as the emission probability of CRF model, and the transfer probability of external influencing factors is applied to the emotion classification task processed by CRF model. The overall frame diagram is shown in Figure 2.

3.1. Text Word Vector Representation

In this paper, use word2vec to train the expectation based on CBOW model, thus obtaining the attached result word vector representation. Word vector can map words in corpus from entity meaning to real number space. The dimension of each word in the specified word vector space is $V$, the size is $|V| \times n$, and is an n-dimensional word vector, which indicates the $|V|$ space size of the word vector, that is, the set of how many words it contains. Then a comment text $T$ in the corpus can be represented as a sequence set as follows: $(x_1, x_2, ..., x_n)$.

The number of words of one piece of news information represented by the above $n$; $x_i$ represents the I word in the news text. When news text is converted into a low-dimensional word vector matrix, the word vector represented by the word should be searched in $V$. Existence means $Z$, otherwise the $Z$ is 0, which can be expressed as follows: $(x_1, x_2, ..., x_n) \Rightarrow (Z_1, Z_2, ..., Z_n).

3.2. BILSTM-CRF Emotion Classification Model

As shown in the overall architecture diagram of the above figure, the BILSTM-CRF model constructed in this paper consists of an input layer, a word embedding layer, a bidirectional LSTM layer and a CRF layer. The output of each layer is the input of the next layer. The depth word vector features of emotion classification can be obtained by using this model. The model is constructed as follows:

1) Input layer. This layer is the input part of the model, that is, the input of a text content in the corpus is processed later.

2) Word embedding layer. The vector list of the context is obtained through the training of the corpus by Word2vec, and then the vector corresponding to the input words is searched in the vector list, and the vector is synthesized into a matrix in the following form:

Figure 2. The Overall Frame Diagram.
The i-th line of $Z$ represents the word vector of the i-th word of news text information, and the dimension is m-dimension.

3) Two-way LSTM layer, which is equivalent to the feature extraction part, LSTM neural network is constructed from both positive and negative directions, which can better capture the contextual semantic information of sentences and better reflect the semantic expression of texts. The input of the two neural networks is consistent.

LSTM has three stages: input stage, forgetting stage and output stage. These three stages adjust and control the input timing information well. The forward calculation process at a certain time is as follows:

Input Gate Principle: 
$$i_t = \sigma(W_i \times [h_{t-1}, x_t] + b_i)$$
$$C_t = \tanh(W_c \times [h_{t-1}, x_t] + b_c)$$
$$f_t = \sigma(W_f \times h_{t-1} + b_f)$$
$$o_t = \sigma(W_o \times h_{t-1} + b_o)$$
$$Z_t = o_t \times \tanh(C_t)$$

Forgetting Gate Principle: 
$$f_t = \sigma(W_f \times h_{t-1} + b_f)$$

Output gate principle: 
$$o_t = \sigma(W_o \times h_{t-1} + b_o)$$

Where $W, b$ is the parameter matrix of the network for each layer; $i, f, o$ and respectively represent the input Matrix in Memory Stage, Output Results in Forgetting Stage, Input and Output Stage; $h_{t-1}, x_t$ which respectively represent the internal states of the memory stage at time t; $Z_t$ Represents the output of the memory stage at time t.

The output of the forward LSTM language model $\tilde{h}$, and the output of the reverse LSTM language model $\hat{h}$ are subjected to vector splicing to obtain feature vectors for emotional analysis:
$$\tilde{h} = \{\tilde{h}_1, \tilde{h}_2, ..., \tilde{h}_n\}$$
$$\hat{h} = \{\hat{h}_1, \hat{h}_2, ..., \hat{h}_n\}$$

This vector is the final feature obtained by BILSTM model, which is used as the basis for generating the transmission probability of CRF classification model in this paper. Firstly, the new feature vector is linearly transformed to obtain the score of each emotion category of the current word, $score = Wh + b$.

Where b is biased, v is emotional villa, and score's dimension is v. The scores of the current sentence belonging to each emotional category are score1, score2,..., scorev respectively.

Use Softmax to calculate the probability of each kind of emotion in each news text:
$$p(y_i | h) = \frac{e^{score_y_i}}{\sum_{j=1}^{v} e^{score_j}}$$

It indicates whether the I-th dimension is the target emotion category of the current input sequence, if the I-th dimension is the target emotion category of public opinion, it is 1, otherwise it is 0, and the calculated probability value is taken as the emission probability of CRF model.

The labeling sequence with the highest probability among all possible sequence labeling results is calculated as the final prediction result of the model. The probability of external influencing factors is taken as the transfer probability, and its formula can be defined as: $p_{ij} = \phi(y_{ij}, y_{ij-1})$

Where $y_{ij}$ is the target emotion category, $y_{ij-1}$ is the category identification of the previous word, $p_{ij}$ and the probability of transferring from a certain emotion category identification I of the previous word to a certain emotion category identification J of the current word are represented by a function $\phi$ to find the transfer probability.

The probability of the result sequence is obtained by multiplying the emission probability and the transition probability of the response. The calculation method for the probability of the result sequence is shown in the following formula:
$$P(Y | X) = \frac{1}{Z} \sum_{Y \in \mathcal{Y}} e^{\sum_{t=1}^{T} \log p(y_t | y_{t-1})}$$

Where $Z$ is the normalizing factor, which is the probability of all possible sequence labeling results.

With the help of the above formula, the class with the largest score is obtained, which is the classification result of the news text.

4. Experimental Results and Analysis

The experimental data come from public opinion news information about novel coronavirus pneumonia on major websites. After deduplication and preprocessing of the data set, 1500 news text information samples of the training set were finally obtained, including 1217 samples with very positive emotions,
157 samples with very negative emotions, 63 samples with relatively negative emotions and 63 samples with relatively positive emotions. The sample of the test set is 600 news samples.

4.1. Evaluation Index
In this paper, the accuracy rate $P$ (Precision), recall rate $R$ (Recall), F-measure and micro-average estimation are selected to evaluate the emotion model of BILSTM and CRF. The calculation formula of micro-average is as follows:

$$P = \frac{\sum \text{correct}(e = i)}{\sum \text{proposed}(e = i)}, \quad R = \frac{\sum \text{correct}(e = i)}{\sum \text{gold}(e = i)}, \quad F = \frac{2 \times P \times R}{P + R}$$

Among them, sys_correct represents the quantification of the consistent results predicted by the model and the target, gold represents the number of public opinion target emotions, sys_proposed represents the number of model labels, and $I$ is the value of 4 types of emotions.

4.2. Comparative Analysis of Experiments
In order to show that the BILSTM-CRF model proposed in this paper is better, this paper selects LSTM, BILSTM and BILSTM-CRF models to carry out the same experiment as the experimental control. It is verified that the model proposed in this paper has stronger generalization ability, as shown in Figure 3 and 4.

As shown in the table, compared with BILSTM, LSTM has poor results. The specific reason is that the semantic relationship between contexts is not well considered and the dependency relationship between semantics cannot be well captured. The changes of F1 value and recall rate during model training are shown in the figure. As can be seen from Figure 3 and 4, the F1 value, recall rate and convergence rate of the BILSTM-CRF model proposed in this paper compared with the comparative experiment, it is greatly improved.

![Figure 3. F1 value change.](image)

![Figure 4. Recall value.](image)

5. Summary and Prospect
Based on the emotion prediction of public opinion energy events, this paper proposes a kind of emotion classification based on traditional meaning, which takes external influencing factors as a state transition matrix and assigns values to CRF's state matrix, so as to more accurately predict the trend of public opinion events emotion analysis. Firstly, word is embedded based on Word2vec to obtain word vector representation, Furthermore, BILSTM-CRF neural network model involved in this paper is used to extract the depth word vector features of the text, and after linear transformation and Softmax function, it is used as the emission probability of CRF model, and the state transition matrix of external influencing factors is added. Finally, the result probability is multiplied by the corresponding emission probability and transition matrix probability to obtain the optimal emotion classification result. The experimental results show that this method can effectively classify the emotion about Energy in network public opinion news. With the development of big data, the emotion classifier of this model will be further optimized for in-depth research.

Acknowledgments
This research was partially supported by the CETC Joint Fund, the National Natural Science Foundation of China (No. 62072089), the China Postdoctoral Science Foundation (Nos. 2019T120216 and
2018M641705), the Fundamental Research Funds for the Central Universities (Nos. N2019007, N180101028, N180408019 and N2024005-2) and the fund of Acoustics Science and Technology Laboratory.

References

[1] Xiaohu Meng, Li Li. Research on the Law of Network Public Opinion Dissemination in Energy Industry [J]. Journalism Research Guide, 2017, 8(03): 144-145.

[2] Lixin Song. Perfect the Public Opinion Disposal Mechanism of Energy Enterprises [J]. Sinopec, 2014(10): 104-105.

[3] Chen Long, Guan Ziyu, He Jinhong, et al. A survey on sentiment classification [J]. Journal of Computer Research and Development, 2017, 54(6): 1150-1170.

[4] Zhang Xiangyang, Na Risa. Emotional classification feature selection based on complex network [J]. Application Research of Computers, 2017, 34(4): 1000-1003.

[5] Li Shoushan, Huang Lei, Wang Rong, et al. Sentence-level emotion classification with label and context dependence [C] // Proc of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing. Stroudsburg, PA : Association for Computational Linguistics, 2015: 1045-1053.

[6] Martineau J, Finin T. Delta TFI DF: An Improved Feature Space for Sentiment Analysis [C] // International Conference on Weblogs and Social Media, ICWSM 2009, San Jose, California, USA, May. DBLP, 2009.

[7] Dongping Wei, Niansheng Tang, Tianli lei, Shouwen Wen. The deep learning word vector model using part of speech and sentiment information [J]. Journal of Intelligent & Fuzzy Systems, 2019, 38(1).

[8] Dong Qiu, Haihuan Jiang, Shuqiao Chen. Fuzzy Information Retrieval Based on Continuous Bag-of-Words Model [J]. Symmetry, 2020, 12(2).

[9] Hao PENG, Jianxin LI, Hao YAN, Qiran GONG, Senzhang WANG, Lin LIU, Lihong WANG, Xiang REN. Dynamic network embedding via incremental skip-gram with negative sampling [J]. Science China (Information Sciences), 2020, 63(10): 89-107.

[10] Zhu Xiaodan, Sobhani P, Guo Hongyu. Long short-term memory over recursive structures [C] // Proc of the 32nd International Conference on International Conference on Machine Learning, New York: ACM Press, 2015: 1604-1612.

[11] Arabic Diacritization Using Bidirectional Long Short-term Memory Neural Networks With Conditional Random Fields [J]. Technology News Focus, 2020.