Received August 27, 2020, accepted September 14, 2020, date of publication September 18, 2020,
date of current version September 30, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3024750

Sentiment Analysis About Investors and Consumers in Energy Market Based on BERT-BiLSTM

REN CAI[1,2], BIN QIN[2], YANGKEN CHEN[1,2], LIANG ZHANG[2,3], RUIJIANG YANG[4], SHIWEI CHEN[1,2], AND WEI WANG[1,2]

[1]College of Mechatronics and Control Engineering, Shenzhen University, Shenzhen 518060, China
[2]Information Center, Shenzhen University, Shenzhen 518060, China
[3]College of Electronics and Information Engineering, Shenzhen University, Shenzhen 518060, China
[4]College of Electrical and Information Engineering, Hunan University, Changsha 410082, China

Corresponding author: Bin Qin (qinbin@szu.edu.cn)

This work was supported in part by the Postgraduate Education Reform Project of Shenzhen University, in 2020, under Grant 860-000001050503, and in part by the Special-purpose Funds for anti-epidemic of Shenzhen University, in 2020.

ABSTRACT With the rapid development of social media, the number of online comments has exploded, and more and more people are willing to express their attitudes and feelings on the Internet. Under the influence of a series of major events all over the world, production order is facing a serious challenge, which severely impact on energy market. In 2020, a large number of investors' and consumers' comments related to social events began to appear on the Internet from China. However, the style and quality of online comments vary greatly, making it difficult to accurately extract users' views and tendencies. Based on the investors' and consumers' statements published on Chinese Internet, in this paper, we use statistical methods to classify the sentiment orientation of the Netizens firstly, and then use Bidirectional Encoder Representations from Transformer-Bidirectional Long Short Term Memory (BERT-BiLSTM) which is the combination forecasting method of Bidirectional Encoder Representations from Transformer (BERT) and Bidirectional Long Short Term Memory (BiLSTM), to model and forecast the sentiment orientation of users' statements, as well as being compared with its based models, BERT and BiLSTM. Among them, the accuracy and recall, which represent the predictive abilities on the overall samples and on the focus(the samples of Label 1) respectively, of BERT-BiLSTM model are 0.8620 and 0.7078 respectively, which are superior to those of BERT model (0.8559 and 0.5576) and LSTM model (0.7775 and 0.0747). The research results can accurately predict the sentiment orientation of Internet users during the social events so as to provide technical support for grasping the energy market trend.

INDEX TERMS Sentiment analysis, combination model, BERT-BiLSTM, energy market, investors and consumers.

I. INTRODUCTION

A. MOTIVATION

Energy is an important pillar for modern social and economic development. In the global energy consumption structure, traditional fossil fuels such as oil, heating oil, gasoline and natural gas account for more than 80% of the global energy share and exert a profound influence on the world economy. Public sentiment, however, has a profound effect on energy markets. In particular, the occurrence of major events will greatly affect the sentiment of investors and consumers, thus affecting energy market. In order to strengthen the management of the energy market, it is necessary to model and predict the public sentiment, so as to provide technical support for the mastery of confidence in the energy market. Between the end of 2019 and the first half of 2020, there occur a series of major events all over the world, such as disease, whose infection way includes droplets and contact transmission [1], [2]. Major public events have derivative characteristics [3]. As the event develops and materials supplies become scarce, public anxiety and panic inevitably appear. Once the trend spread and aggravate, it will seriously disrupt the

The associate editor coordinating the review of this manuscript and approving it for publication was Junjie Hu.

production order and cause potential impact on energy market. The fluctuation of energy market is large, which cannot be completely explained by the traditional financial theory. According to academic and industry studies, investors' and consumers' sentiments have an impact on the volatility of the energy futures market. Speculative sentiment in the energy futures market is more influenced by good news than bad news, as speculators more inclined to believe in a positive market outlook. The research of Chen *et al.* shows that there is a significant negative relationship between the changes of Internet investors' sentiment and market returns [4]. Antonakakis *et al.* study the dynamic relationship between economic policy uncertainty index and change in oil price, indicating that economic policy uncertainty impacts the investors' and customers' sentiments, which have a negative response to the oil price shock of aggregate demand [5].

In order to grasp the energy market trend, it is necessary to carry out sentiment analysis for the investors and consumers in energy market.

B. LITERATURE REVIEW

We would discuss the popular methods that are used in among academic and industry community about sentiment analysis in recent years. According to their characteristic, these methods could be classified as the following:

1) RULE-BASED METHOD

Rule-based method relies on 'dictionary + rule' mode. Turney [6] points out that contain the phrase of the adjectives and adverbs is judgement sentence emotional tendency of important basis. Thus, he puts forward Point Mutual Information (PMI) to calculate the emotional value of vocabulary.

In recent years, the rise of social networking, makes the emoticons become a main way for people to express their emotions. So Hamouda *et al.* propose a contains emoticons emotional vocabulary for emotion recognition [7]; Gaikwad *et al.* find that the contributions of different words and symbols to sentence emotional polarity are different so that the weight should be put forward to distinguish the influence of these words [8]. The method relies on the building of dictionary and selection, is poor in adaptability for different fields and can't solve the problem of complex ambiguity word by itself. Such method usually functions as an auxiliary method in combination with other methods to be applied to practice.

2) MACHINE LEARNING METHOD

The core of machine learning method lies in the selection of effective feature combinations and the use of classifiers to classify emotions. Pang *et al*. apply machine learning algorithm to emotion classification task for the first time. By comparing the performance of Naive Bayes (NB), Maximum Entropy (ME) and Support Vector Machine (SVM) combined with different features on emotion classification, they find that the accuracy of SVM combined with UNIGRAMS features could reach 82.9% [9]. Zou and Tang *et al.* take into account the complex syntactic relationships in the text, which combine word positions, dependence or grammatical relations between words and part-of speech characteristics, effectively improving the emotional classification performance [10]. Although machine learning method achieves good results, it requires a large number of manual annotation data sets to train the model, and the high-quality data sets cost a lot of time and labor. The emotion classification performance of machine learning method is also limited by the complex feature design, and it has poor adaptability in different fields. Therefore, finding effective feature combination is its future development direction.

3) RULE-BASED METHOD + MACHINE LEARNING METHOD

To solve the above problems, some scholars propose to combine dictionary and machine learning to classify emotions by combining rules as features with other features and using classifiers to classify [11]. But it is still not out of the constraints of dictionary and complex feature engineering.

4) WEAK-TAGGING-BASED METHOD

In machine learning method, it costs a lot of time and manpower to obtain high quality manual annotation data sets. Some scholars proposed to mine the weakly marked information reflecting emotional semantics in the comment data to a certain extent, such as the rating information in the microblog emoticons [12], etc., to train the emotion classifier. However, since this method fails to eliminate the noise influence in the weakly labeled information and takes the weakly labeled information as the feature of emotion classification, it fails to capture the complex feature function from text to deep semantics. The emergence of deep learning solves this problem.

5) DEEP LEARNING METHOD

Deep neural network can simulate the hierarchical structure of human brain and make it have deeper emotional semantic representation ability, so it has been widely used in sentiment analysis. Some scholars have established a series of emotion classification models based on neural network, including a classification model based on convolutional neural network [13], a classification model based on long-term and short-term memory [14], and a classification model based on deep belief network [15]. Compared with machine learning, deep learning has got rid of the constraints of complex feature engineering, but supervised deep learning still needs a large number of annotated data set training models, while unsupervised deep learning requires strict semantic correlation in data, so this method still needs to be further developed. At present, semi-supervised learning, which attracts much attention, only needs to use a small amount of labeled data to train the model, which is superior to the above two methods in terms of emotion classification. In addition, the weak annotation information contains certain semantic relevance, and it is also one of the future development directions to use it as the training data of the deep learning model. To sum up, the traditional sentiment analysis method is suitable for solving the problem of emotion classification in small-scale texts. In the face of massive data, the analysis efficiency is low and it is difficult to locate the emotion information quickly. In addition, the traditional sentiment analysis method only judges the emotional tendency of the whole text and fails to analyze the emotion of the deep semantic role. However, the emergence of theme extraction technology solves these problems.

II. ALGORITHMS FOR SENTIMENT ANALYSIS

In this paper, we adopt a combination forecasting model, Bidirectional Encoder Representations from Transformer-Bidirectional Long Short Term Memory (BERT-BiLSTM) to construct the models and predict the sentiment orientation about investors and consumers in energy market. BERT-BiLSTM is an improved algorithm based on Bidirectional Encoder Representations from Transformer (BERT) and Bidirectional Long Short Term Memory (BiLSTM). It's worth noting that BERT-BiLSTM adopts BERT as its upstream part and BiLSTM as its downstream part INSTEAD OF the simple weight-combination model between BERT and BiLSTM.

To specifically prove the advantages of the combination model, BERT-BiLSTM, in the sentiment analysis for energy market, we take use of its based models, BERT and BiLSTM, as the baseline models rather than other models.

Therefore, we will introduce these algorithms in this section. Particularly, for purpose of helping readers understand the principle of BERT-BiLSTM, we will firstly introduce their based models, BERT and BiLSTM in detail, and then brief readers on BERT-BiLSTM.

A. BERT

BERT has 6 layers of Transformer superimposed on the encoder and decoder respectively, which accounts for the extremely complex training process, high configuration, a large amount of training time and very expensive cost. However, Google opens the source of the pretrained model of BERT, which can be directly used for us. In this paper, we use a simplified Chinese training model of Chinese characters. With a total of 12 layer, the model 768 hidden units, 12 since the note head, 110 million parameters. It is to obtain a high-quality characteristics of the word vector by using BERT for the downstream model input [16]. The traditional model of natural language needs to input in the form of digital vector, which usually means need to convert vocabulary and parts of speech to digital characteristics In the past, words were expressed as one-hot encoding or, more useful, embedded as neural words, in which words are matched with feature inserts of fixed length, which are generated by models such as word2vec or GloVe [17]. The overall framework of the BERT model is shown in Fig. 1. BERT offers an advantage over the traditional models such as word2vec. Although each word in word2vec has a fixed representation, that representation is independent of the context in which the word appears.

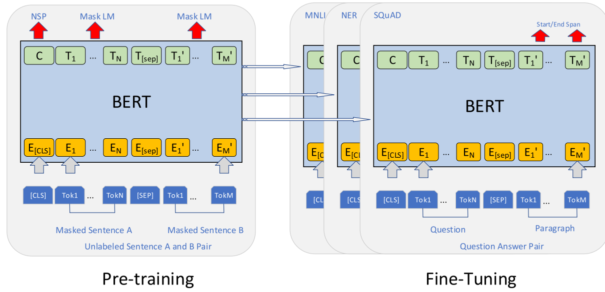

FIGURE 1. The overall framework of BERT.

Comparatively speaking, the representation generated by BERT is dynamically notified by the words around the word. For different natural language processing tasks, the model input will be fine-tuned. For text classification tasks, BERT model inserts a [CLS] symbol before the text, and uses the output vector corresponding to the symbol as the semantic representation of the whole text for text classification. It can be understood that compared with other words or words existing in the text, this symbol without obvious semantic information will more 'fairly' fuse the semantic information of each word or word in the text.

In fact, BERT is a language model. Language model usually takes on a large scale, has nothing to do with a particular natural language processing tasks of training text corpus, whose goal is to learn the language itself should be what kind of. It's like when we study Chinese, English and other language courses, which require us to learn how to choose the words already known to produce a coherent text. The pre-training process of BERT model is to gradually adjust the model parameters, so that the semantic representation of text output by the model can describe the essence of the language and facilitate the subsequent fine-tuning for specific NLP tasks. To achieve this goal, BERT proposes two pre-training tasks: Masked Language Modeling (Masked LM) and Next Sentence.

The mission of Masked LM is to, for a given sentence, randomly masked one or several words of this sentence to ask you to predict what the masked words are according to the remaining words, which is similar to gestalt. The pre-training process of BERT model is essentially mimicking our language learning process. Specifically, 15% of the words in a sentence were randomly selected for prediction. More precisely, the words that have been erased from the original sentence, 80%, 10% and 10% of which are replaced by special [MASK], arbitrary words and themselves respectively. The main reason for this is that the [MASK] marker will not be used in the subsequent fine-tuning tasks. Another advantage of this is that when predicting a word, the model does not know whether the input word in the corresponding position is the correct word, which forces the model to rely more on the context information to predict the word, which gives the model a certain error correction ability.

The task of Next Sentence Prediction is to judge whether the second sentence in the text follows after the first sentence in two given sentences of an article, similar to the paragraphs reordering in the English exam. Specifically speaking, it is to mess up the paragraphs of the passage, and then let us reorder them, which requires us to have full and accurate understanding. The Next Sentence Prediction is essentially a simplified version of paragraph reordering: considering only two sentences and determining if it's a previous or next sentence in an article. In the actual pre-training process, 50% correct statement pairs and 50% wrong statement pairs are randomly selected from the text corpus for training.

It allows the model to more accurately characterize the semantic information at the declarative and even textual level through combining Next Sentence Prediction with the Masked LM. The BERT model performs joint training on the Masked LM and Next Sentence Prediction, so that the output vector representation of the model can depict the overall information of the input text as fully and accurately as possible, providing a better initial value of model parameters for subsequent fine-tuning tasks.

Transformer is the core module of BERT, and the attention mechanism is the most critical part of Transformer. The main function of the attention mechanism is to get the neural network to focus "attention" on one part of the input, that is, to distinguish the effects of different parts of the input on the output. Attention mechanism can enhance semantic representation. The meaning expressed by a word in a text is usually related to its context, and the context information of a word helps to enhance its semantic representation. At the same time, different words in the context often play different roles in enhancing semantic representation. In order to enhance the semantic representation of the target word by discriminating the context information, attention mechanism can be used. The self-attention mechanism is to enhance the semantic vector representation of each word in the input text. And Multi-head Self-Attention is to enhance the diversity of Attention mechanism, to increase the use of different Attention mechanism module to obtain each word in the text under different semantic space enhancement semantic vector, and every word multiple enhancing semantic vector linear combination, in a vector of the same length as the original word semantic vector. Transformer, on the other hand, adds three key operations to the Multi-head Self-Attention mechanism, namely residual connection, standardization, and linear transformation. Residual Connection means that the input and output of the module are added directly as the final output. A basic consideration behind this operation is that it is easier to modify the input than to reconstruct the entire output, thus making the network easier to be trained. The Normalization refers to the Normalization of the mean to 0 and variance to 1 of one Layer of neural network nodes.

Linear transformation means that two linear transformations are performed on the enhanced semantic vector of each word to enhance the expression ability of the whole model, and the transformed vector remains the same length as the original vector.

B. LSTM

BiLSTM model BiLSTM is composed of bidirectional LSTM. LSTM is an improved model of RNN, which solves the problem of derivative explosion and derivative loss existing in RNN, and can process sequence data of different length, thus overcoming the problem of rapid loss of information in circulating neurons [18].

LSTM is made up of the input word x_t at the moment, Cell state c_t, temporary cell state c_t, hidden layer state h_t, forgetting gate f_t, memory gate i_t, output gate o_t. LSTM has three main stages: forgetting stage, selective memory stage and output stage. Firstly, the forgetting stage, in which information in cells is selectively forgotten and important information is retained. This stage is controlled by the forgetting gate $\mathrm{f_t}$. Secondly, it comes the selective memory stage, in which information in the input cells is selectively "remembered". Important information is emphasized and unimportant information is discarded. This stage is controlled by the memory gate i_t. Finally, it is the output stage, which determines which information will be selected as the output, controlled by the output gate to. The overall framework of the LSTM model is shown in Fig. 2.

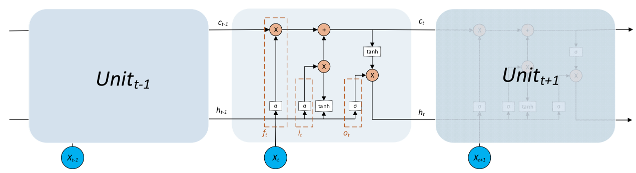

FIGURE 2. **The overall framework of the LSTM.**

BiLSTM consists of two parts, namely, the forward LSTM and the backward LSTM. Different from the forward LSTM, the backward LSTM is to reverse the input sequence and calculate the output in the way of the forward LSTM. After simply stacking the output of the forward LSTM and the output of the backward LSTM, the final result of the BiLSTM model is obtained, enabling the model to consider both the above information and the following information.

C. BERT-BiLSTM

Based on BERT and BiLSTM, a method combining the above two methods, BERT-BiLSTM, is be applied to construct the model and predict the sentiment orientation about investors and consumers in energy market in this paper, whose structure is shown in the figure below:

Instead of the simple weight-combination model between BERT and BiLSTM, BERT-BiLSTM adopts BERT as its upstream part and BiLSTM as its downstream part. This is because according to the introduction above, BERT has the ability to learn the statistics characteristic of the nearby words and BiLSTM is competent in learning the context information, which accord with logic of the human language system that the basic grammar depends on the statistical characteristics and the specific meaning hinges on the context. Thus, BERT-BiLSTM has the potential to analyze the

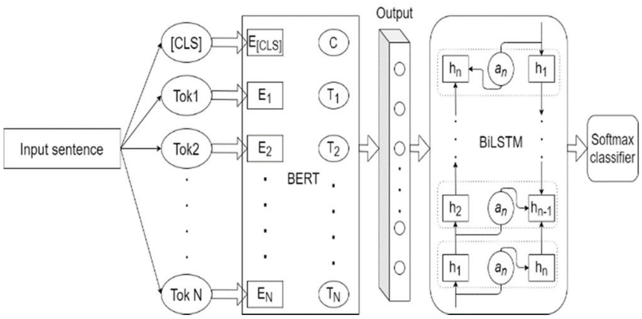

FIGURE 3. The structure of BERT-BiLSTM.

sentiment about investors and consumers in energy market with reference to their comments in the Internet.

The model takes the output $\boldsymbol{C} \in \boldsymbol{R}^{n}$ of the last layer trained in BERT including the [CLS], and add the weight $\boldsymbol{W_a} \in \boldsymbol{R}^{d_a \times n}$ as the input of the BiLSTM model, the calculation is as shown in the formula:

$$a_i = g_1 \left(W_a C_i + b_a \right) \tag{1}$$

where $1 \leq i \leq$ n, n is the dimension of feature vector of a sentence after BERT training; $\boldsymbol{a_i} \in \boldsymbol{R}^{d_a}$; $\boldsymbol{b_a}$ is the bias vector with the dimension $\boldsymbol{d_a}$; the activation function $\boldsymbol{g_1}$ adopts the Sigmoid function.

The model inputs the input vector into the hidden layer. A standard LSTM calculates the hidden layer vector sequence $\boldsymbol{h}$ from one direction, while the BiLSTM calculates two hidden layers in different directions, and finally combines the results of the two directions to output. The forward hidden layer vector and the backward hidden layer vector are $\vec{\boldsymbol{h}}$ and $\overleftarrow{\boldsymbol{h}}$ respectively, and the output vector of the hidden layer at the i-th moment is $\boldsymbol{v_i}$. The calculation method is as follows:

$$v_i = \vec{h}_i + \overleftarrow{h}_i \tag{2}$$

where $\vec{\boldsymbol{h}}_i \in \boldsymbol{R}^{d_h}$, $\overleftarrow{\boldsymbol{h}}_i \in \boldsymbol{R}^{d_h}$.

In addition, the model takes use of tanh function as the activation function $\boldsymbol{g_2}$ to calculate the hidden layer, where $\boldsymbol{h}$ is calculated as shown in the formula:

$$h_i^d = g_2 \left(W_h^d a_i + U h_{i-1}^d + b_h^d \right) \tag{3}$$

where $\boldsymbol{W_h^d} \in \boldsymbol{R}^{d_h \times d_a}$ is the weight matrix of $\boldsymbol{a_i}$, corresponding to the d-th index; $\boldsymbol{U}$ is the corresponding weight matrix of the hidden layer output $\boldsymbol{h^d}$ at time i-1; $\boldsymbol{d} \in \{0, 1\}$ represents two different directions in the hidden layer; $\boldsymbol{b_h^d} \in \boldsymbol{R}^{d_h}$ is the bias vector corresponding to d-th index.

Then all the hidden layer $\boldsymbol{h_i^d}$ are connected and combined into a vector $\boldsymbol{H}$, which is the final sentence-level feature vector. Then input the feature vector $\boldsymbol{H}$ to the fully connected layer, and use ReLU function as the activation function. Finally, the output of the fully connected layer is used as the input of the output layer, and the Softmax function is used for classification. The probability calculation of the final sentiment orientation classification is shown in the formula:

$$p \left(y | H, W_s, b_s \right) = \text{softmax} \left(W_s H + b_s \right) \tag{4}$$

where $\boldsymbol{W_s} \in \boldsymbol{R}^{|s| \times |l|}$ and $\boldsymbol{b_s} \in \boldsymbol{R}^{|l|}$ are the parameters of the output layer; $|l|$ is the number of categories.

III. TRAINING METHOD AND PROCESS FOR SENTIMENT ANALYSIS

The abstract method and process of training for BERT-BiLSTM is shown as Fig. 4:

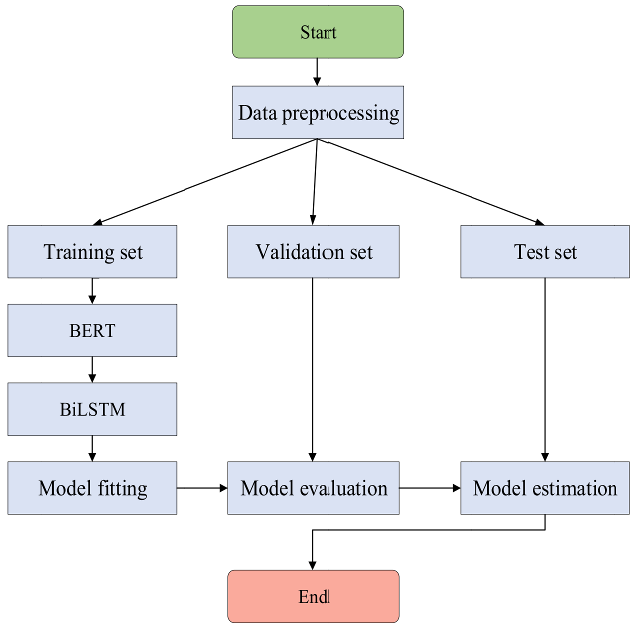

FIGURE 4. The process of training of BERT-BiLSTM.

The specific method and training steps in this paper can be summarized as five steps:

1) Firstly, preprocess the data to eliminate invalid data in the original data (such as empty text, text with only punctuation, etc.).

2) Secondly, the data set is divided into training set, validation set and test set.

3) Thirdly, the data set is sent to BERT for training, and get the preliminary word vector.

4) Fourthly, the preliminary word vector is sent into BiLSTM.

5) Finally, the output of BiLSTM is sent to the Sotfmax layer to get the final sentiment orientation classification matrix.

It's worth nothing that in Fig. 4, 'BERT' and 'BiLSTM' are the upstream part and downstream part of BERT-BiLSTM respectively. In addition, the training methods and processes of BERT and BiLSTM are similar.

IV. PERFORMANCE METRICS

There are many kinds of evaluation indicators for the two-category model. Thus, choosing the right evaluation indicators based on the problem is an important method to measure the good or bad. We adopt the following methods commonly used in the academic circles [19], [20].

A. ACCURACY

Accuracy is the simplest evaluation indicator in classification problems, but there are fatal defects in the training set with unbalanced samples. For example, when the negative sample ratio is 99%, the model only predicts the result as for negative samples, the accuracy rate is 99%, but this is not the most reasonable way to evaluate.

$$accuracy = \frac{n_{correct}}{n_{total}} \tag{5}$$

where $n_{correct}$ represents the number of samples that are accurately classified in the test set, and n_{total} represents the total number of samples in the test set.

B. PRECISION AND RECALL

Precision and recall are a pair of indicators for measuring the retrieval system. The accuracy is the ratio of the correct result pushed by the retrieval system to the total results pushed. The recall is measured by the system. The correct result is proportional to the actual correct result of the entire system.

Mapping to this problem, the accuracy indicates how much of the positive samples (investors and consumers in energy market of Label 1) predicted by the model is the true positive samples. The higher the value, the stronger the ability of the model to correctly identify the true positive samples. The recall indicates the ratio of the true positive samples identified in all of the positive samples. The higher the value, the more true positive samples can be identified.

However, the accuracy and the recall are a pair of contradictory indicators. In order to improve the accuracy rate, the model needs to be more certain to identify the true positive sample, which will inevitably give up some of the many not grasped positive sample (but actually the positive sample), resulting in a lower recall, the relationship about precision and recall are shown as table 1 and formula 6 to 7.

$$Precision = \frac{TP}{TP + FP} \tag{6}$$

$$Recall = \frac{TP}{TP + FN} \tag{7}$$

TABLE 1. Confusion matrix.

	positive sample in forecast	negative sample in forecast
positive sample in reality	TP	FN
negative sample in reality	FP	TN

C. F1 SCORE

In order to more comprehensively evaluate the model, the F_1 *Score* [21] is introduced. The F_1 Score is the harmonic mean of the precision rate and the recall rate:

$$F_1 Score = \frac{2 \times precision \times recall}{precision + recall} \tag{8}$$

V. CASE STUDY

In this paper, we collect about 100 thousand investors' and consumers' comments samples published on Chinese Internet as the source dataset. Parts of the samples are shown as table 2.

TABLE 2. Parts of the samples.

Sample ID	The Original Comments (The English Meaning of them)	Label
000001	近期，因全球公共卫生事件等不确定性因素对于全球原油需求造成了负面影响。(Recently, uncertainties such as global public health events have had a negative impact on global oil demand.)	0
000002	工厂、工地推迟复工。(Factories and construction sites have been delayed in their resumption.)	1
000003	油价周四攀升，因投资者希望全球最大的产油国将同意进一步减产(Oil prices rose on Thursday because the investors hope that the world's largest producer would agree to further production cuts.)	0
000004	氢能汽车开启商业推广。(Hydrogen cars are come into commercial popularizing.)	0
…	…	…

According to the training method and process in Section III, the data set is randomly divided into three sub data sets: training set, validation set and test set, with 7:1:2 scale.

Then, after upsampling and downsampling for training set, the ratio of positive samples (negative sentiment samples) to negative samples (positive and neutral sentiment samples) is 1:2, which is more balanced than that in source data set, about 1:5.

Next, all of the sub data sets are normalized and used to train and test the model.

The result are shown in the following figures and table.

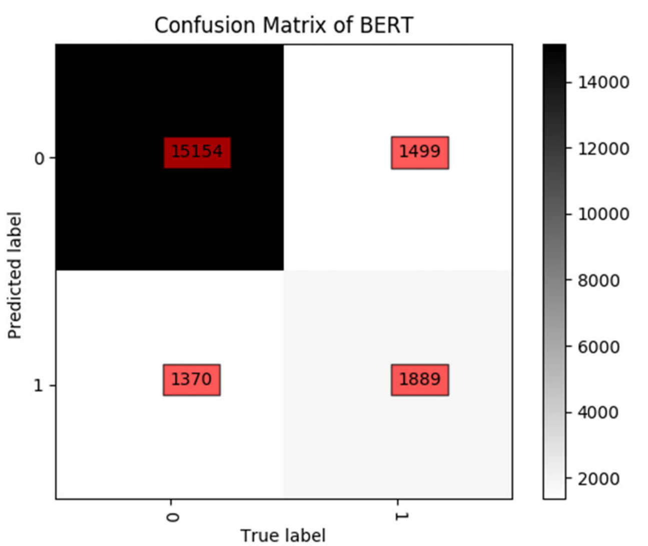

FIGURE 5. Confusion matrix of BERT.

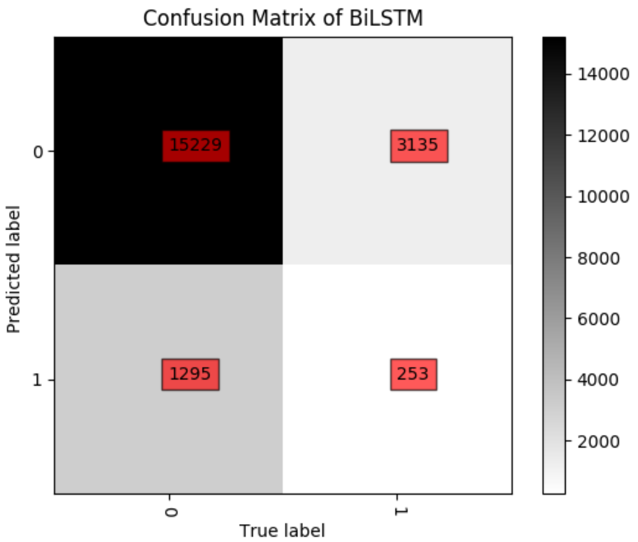

FIGURE 6. Confusion matrix of BiLSTM.

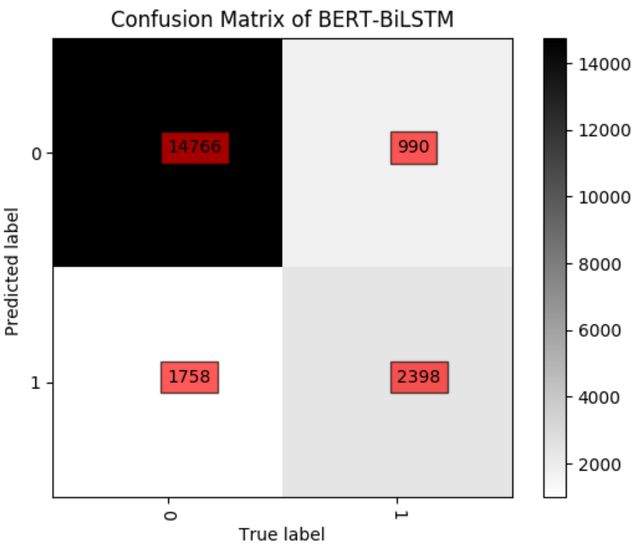

FIGURE 7. Confusion matrix of BERT-BiLSTM.

The table 3 shows the metrics of the BERT, BiLSTM and BERT-BiLSTM respectively.

TABLE 3. Metrics of the models.

Model	Recall	Accuracy	Precision	F_1 Score
BERT	0.5576	0.8559	0.5796	0.5684
BiLSTM	0.0747	0.7775	0.1634	0.1025
BERT-BiLSTM	0.7078	0.8620	0.5770	0.6357

On the whole, BERT-BiLSTM and BERT models get preferable accuracy (0.8620 and 0.8559 respectively) to the one of LSTM model, so do they get more excellent on other metrics.

Further, since the most important target is to seek the samples of sentiment orientation in Label 1, so that we focus more on the recall of them. Beyond all doubt, BERT-BiLSTM model get a better performance (0.7078) than other two models in this metric. The Intuitive representation is that from Fig. 5 to Fig. 7, BERT-BiLSTM, BERT and BiLSTM find out 2398, 1889 and 253 in 3388 samples of Label 1 respectively, which reveal the superiority of BERT-BiLSTM.

Above all, we use BERT-BiLSTM model to predict the sentiment orientation.

TABLE 4. Parts of the forecast samples.

The Original Comments (The English Meaning of them)	Original Label	Forecast Label
石油输出国组织在其最新月度报告中下调了 2020 年的全球石油需求量预期。(In the latest monthly report, the Organization of the Petroleum Exporting Countries lowered its forecast for global oil demand in 2020)	0	0
搞不明白为什么有人还不支持复产复工。跟我们关系一点没有。(I don't know why some people still don't support the resumption of production and work. The event has nothing to do with us.)	1	1
…	…	…

VI. CONCLUSION

In this paper, the model of BERT, LSTM and BERT-LSTM are applied to forecast the sentiment orientation. Among these models, BERT-LSTM gets the highest accuracy and recall scores, which represent the predictive abilities on the overall samples and on the focus respectively.

The result reveals the advantages of the combination forecasting model, BERT-BiLSTM on sentiment analysis. which can accurately predict the sentiment orientation of Internet users during the major events so as to provide technical support for the decision-making of energy market.

However, it's a pity that we could not get more data and has no enough computing power. Therefore, we would fight for a larger dataset and a better computing power, which may help our model get a more excellent performance.

REFERENCES

[1] C. Huang *et al.*, "Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China," *Lancet*, vol. 395, no. 10223, pp. 497–506, Feb. 2020, doi: 10.1016/S0140-6736(20)30183-5.

[2] C. Wang, P. W. Horby, F. G. Hayden, and G. F. Gao, "A novel coronavirus outbreak of global health concern," *Lancet*, vol. 395, no. 10223, pp. 473–490, Feb. 2020.

[3] World Health Organization. (Jan. 22, 2019). *Ten threats to global health in 2019*. [Online]. Available: https: // www. who.int/emergencies/ten-threats-to-global-health-in-2019

[4] R. Chen, J. Yu, C. Jin, and W. Bao, "Internet finance investor sentiment and return comovement," *Pacific-Basin Finance J.*, vol. 56, pp. 151–161, Sep. 2019.

[5] N. Antonakakis, I. Chatziantoniou, and G. Filis, "Dynamic spillovers of oil price shocks and economic policy uncertainty," *Energy Econ.*, vol. 44, pp. 433–447, Jul. 2014.

[6] P. D. Turney, "Thumbs up or thumbs down?: Semantic orientation applied to unsupervised classification of reviews," in *Proc. 40th Annu. Meeting Assoc. Comput. Linguistics ACL*, 2001, pp. 1–8, doi: 10.3115/1073083.1073153.

[7] S. B. Hamouda and J. Akaichi, ‘‘Social networks text mining for sentiment classification: The case of Facebook’ statuses updates in the, ‘Arabic spring’?Era,’’ *Int. J. Appl. Innov. Eng. Manage.*, vol. 2, no. 5, pp. 470–478 May 2013.

[8] G. Gaikwad and D. J. Joshi, ‘‘Multiclass mood classification on Twitter using lexicon dictionary and machine learning algorithms,’’ in *Proc. Int. Conf. Inventive Comput. Technol. (ICICT)*, Aug. 2016, pp. 1–6. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/7823247/

[9] B. Pang, L. Lee, and V. Shivakumar, ‘‘Thumbs up? Sentiment classification using machine learning techniques,’’ in *Proc. Conf. Empirical Methods Natural Lang. Process. (EMNLP)*, Jul. 2002, pp. 79–86, doi: 10.3115/1118693.1118704.

[10] H. Zou, X. Tang, B. Xie, and B. Liu, ‘‘Sentiment classification using machine learning techniques with syntax features,’’ in *Proc. Int. Conf. Comput. Sci. Comput. Intell. (CSCI)*, Dec. 2015, pp. 175–179. [Online]. Available: http://ieeexplore.ieee.org/document/7424086/

[11] J. Jiang and R. Xia, ‘‘Microblog sentiment classification via combining rule-based and machine learning methods,’’ *Acta Scientiarum Naturalium Universitatis Pekinensis*, vol. 53, no. 2, pp. 247–254 May 2017.

[12] D. Tang, B. Qin, and T. Liu, ‘‘Deep learning for sentiment analysis: Successful approaches and future challenges,’’ *Wiley Interdiscipl. Reviews, Data Mining Knowl. Discovery*, vol. 5, no. 6, pp. 292–303, Nov. 2015.

[13] Y. Kim, ‘‘Convolutional neural networks for sentence classification,’’ 2014, *arXiv:1408.5882*. [Online]. Available: http://arxiv.org/abs/1408.5882

[14] X. Zhu, P. Sobhanip, and H. Guo, ‘‘Long short-term memory over recursive structures,’’ in *Proc. 32nd Int. Conf. Int. Conf. Mach. Learn.*, 2015, pp. 1604–1612, doi: 10.5555/3045118.3045289.

[15] P. Ruangkanokmas, T. Achalakul, and K. Akkarajitsakul, ‘‘Deep belief networks with feature selection for sentiment classification,’’ in *Proc. 7th Int. Conf. Intell. Syst., Modeling Simulation (ISMS)*, Jan. 2016, pp. 9–14. [Online]. Available: https://ieeexplore.ieee.org/document/7877180/

[16] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, ‘‘BERT: Pre-training of deep bidirectional transformers for language understanding,’’ 2018, *arXiv:1810.04805*. [Online]. Available: http://arxiv.org/abs/1810.04805

[17] J. Pennington, R. Socher, and C. Manning, ‘‘Glove: Global vectors for word representation,’’ in *Proc. Conf. Empirical Methods Natural Lang. Process. (EMNLP)*, 2014, pp. 1532–1543. [Online]. Available: https://www.researchgate.net/publication/284576917_Glove_Global_Vectors_for_Word_Representation

[18] S. Hochreiter, ‘‘The vanishing gradient problem during learning recurrent neural nets and problem solutions,’’ *Int. J. Uncertainty, Fuzziness Knowl.-Based Syst.*, vol. 06, no. 02, pp. 107–116, Apr. 1998.

[19] P. Zhang and W. Su, ‘‘Statistical inference on recall, precision and average precision under random selection,’’ in *Proc. 9th Int. Conf. Fuzzy Syst. Knowl. Discovery*, May 2012, pp. 1348–1352. [Online]. Available: https://ieeexplore.ieee.org/document/6234049/

[20] K. H. Brodersen, C. S. Ong, K. E. Stephan, and J. M. Buhmann, ‘‘The binormal assumption on precision-recall curves,’’ in *Proc. 20th Int. Conf. Pattern Recognit.*, Aug. 2010, pp. 4263–4266. [Online]. Available: https://ieeexplore.ieee.org/document/5597760/

[21] X. He, B. D. Gallas, and E. C. Frey, ‘‘Three-class ROC analysis—Toward a general decision theoretic solution,’’ *IEEE Trans. Med. Imag.*, vol. 29, no. 1, pp. 206–215, Jan. 2010.

REN CAI was born in 1994. He is currently pursuing the master’s degree in control science and engineering with Shenzhen University. His research interests include natural language processing, interpretable deep learning, data mining, and high-performance computing.

BIN QIN was born in 1972. He received the bachelor’s degree in software from Xi’an Jiaotong University, in 1993, and the M.Sc. degree in radio electronics from the Xi’an Institute of Optics and Precision Mechanics, Chinese Academy of Sciences, in 1996. He is currently a Senior Engineer with the Information Center, Shenzhen University, and a master’s degree Supervisor of Big Data Cyberspace Security. His research interests include big data information processing technology and artificial intelligence.

YANGKEN CHEN was born in 1997. He is currently pursuing the master’s degree in engineering with Shenzhen University. His research interests include natural language processing, particularly deep learning-based NLP, data mining, and machine learning.

LIANG ZHANG was born in 1996. He is currently pursuing the master’s degree in engineering with Shenzhen University. His research interests include natural language processing, particularly in named entity recognition and information extraction.

RUIJIANG YANG received the B.S. degree in agricultural electrification from Nanjing Agricultural University, Nanjing, China, in 2019. She is currently pursuing the M.S. degree with the College of Electrical and Information Engineering, Hunan University, Changsha, China. Her major research interests include power system planning and operation.

SHIWEI CHEN was born in 1993. He is currently pursuing the master’s degree in control science and engineering with Shenzhen University. His current research interests include recommendation algorithm, machine learning, and natural language processing.

WEI WANG was born in 1992. He is currently pursuing the master’s degree in control science and engineering with Shenzhen University. His research interests include natural language processing and deep learning.