Research on Chinese Text Feature Extraction and Sentiment Analysis Based on Combination Network

Haoyue Xu, Lianhe Yang*

School of Computer Science and Technology, Tianjin Polytechnic University, Tianjin, China
Email: *812157316@qq.com

Abstract

The complexity of Chinese language system brings great challenge to sentiment analysis. Traditional artificial feature selection is easy to cause the problem of inaccurate segmentation semantics. High quality preprocessing results are of great significance to the subsequent network model learning. In order to effectively extract key features of sentences, retain feature words while removing irrelevant noise and reducing vector dimensions, an algorithm module based on sentiment lexicon combined with Word2vec incremental training is proposed in terms of feature engineering. Firstly, the data set is cleaned, and the sentence is segmented by loading a custom sentiment lexicon with Jieba. Secondly, the results after stopping words are obtained through Skip-gram training algorithm to obtain the word vector model. Secondly, the model is added to a large corpus for incremental training to obtain a more accurate word vector model. Finally, the features are learned and classified by inputting the embedding layer into the neural network model. Through the comparison experiment of multiple models, it is found that the combined model (CNN-BiLSTM-Attention) has better classification effect and better application ability.

Subject Areas

Linguistics

Keywords

Feature Selection, Word Vector Representation, Neural Network, CNN-BiLSTM-Attention, Sentiment Analysis

1. Introduction

With the advent of the 5G era, more and more devices will be connected to the Internet, and the resulting data such as text, images, video and audio will see explosive growth. In terms of natural language processing, the proliferation of subjective text resources provides sufficient corpus for emotion analysis. According to a statistical report by The China Internet Network Information Center (2020), China had 904 million Internet users by March 2020, including 710 million online shopping users [1]. Now affected by the epidemic, we seem to have entered an era of "national e-commerce", behind every transaction will be attached to the evaluation of product information, people dig these emotional comments, not only help consumers' shopping decisions, but also for the operators to collect effective feedback information, the shopping experience for the users, merchants, retain customers and the steady development of the society to build economic environment has a major role.

Sentiment analysis [2] is a task of polarity identification for a given content. The traditional methods of text sentiment analysis mainly include the classification method based on sentiment lexicon and the classification method based on machine learning. The dictionary-based method calculates the sentence score by comparing the emotional words in the document with the lexicon and combining a series of rules according to the found results, and finally determines the emotional polarity according to the score [3]. Although the approach based on sentiment lexicon is easy to understand, due to the need for manual annotation, excessive lexicon construction, slightly mechanization, and limited improvement in the performance of the model, the simple use of the dictionary has become rare.

Using machine learning method widely attention in recent years, Meylon *et al.* [4] using Naive Bayesian presidential candidate of the positive and negative comments on Twitter is analyzed, and by comparing the Naive Bayes, K Neighbor and Support Vector Machine classifier's accuracy, drawing a Naive Bayesian classification effect is best, the accuracy is 80.90%. Although shallow machine learning can solve classification problems, compared with deep learning network models, its classification accuracy is limited and cannot achieve good application ability. Therefore, in recent years, various deep learning network models have been widely concerned by researchers.

With the extensive application of deep learning in Natural Language Processing, deep learning algorithms are increasingly used in emotion analysis, and Recurrent Neural Network (RNN) and Convolutional Neural Network (CNN) [5] [6] have become the most commonly used architectural methods. Alayba *et al.* [7] proposed that the CNN combined with the Long and short time memory network (LSTM) adopts the character-level, word-level and Ch5gram level sentiment classification method in different Arabic data sets, and achieved good results. Hanxue Ji *et al.* [8] proposed an sentiment analysis framework of semi-supervised Attention mechanism based on LSTM, which is composed of an

unsupervised Attention LSTM codec and an additional attention-based LSTM model based on Softmax layer. The ability of this algorithm to the task of sentiment analysis is verified by the experimental study of common data sets.

In the essay classification, the traditional vector space model has the problems of feature sparsity and dimensional disaster, and researchers at home and abroad have given a good solution from the feature extension level. Feature extensions include topic model extensions and external knowledge base extensions such as HowNet, WordNet, and so on. Later, it was not until Tomas Mikolov *et al.* [9] released the Word2vec algorithm that this successful word embedding scheme began to be widely used.

The main contributions of this paper are as follows:

1) In terms of acquiring feature words and word vectors, this paper adopts jieba custom external knowledge base for word segmentation to ensure the accuracy of word segmentation; the vector representation of words is obtained by means of Word2vec incremental training. By analyzing and comparing the experimental results, it can be concluded that SL-W2V-Plus, a feature selection method based on sentiment lexicon segmentation and Word2vec incremental training, can obtain higher F1 value and accuracy in the classification model.

2) By learning the mainstream neural network, we know that CNN is a feed forward neural network, which has a strong local learning ability, but cannot transmit information from one layer to another. RNN is good at modeling sequences, but simple RNN cannot solve the problem of gradient disappearance or gradient explosion. Considering the advantages of the two models, a network model method of introducing attention mechanism into the CNN combined bidirectional LSTM is proposed. The experimental results show that the combined model has better performance in feature learning and classification than the baseline model.

2. Improved Feature Engineering

2.1. Feature Extraction

In Chinese text sentiment classification, data preprocessing, as the first step of the whole task, is of great significance. High quality preprocessing results can be better used in model training and testing to improve the accuracy of experimental results analysis. This paper deals with the data in the order of data cleaning, text segmentation and stop words.

First of all, in the process of data cleaning, this paper mainly carries out the following processing on the data: 1) English and Numbers in the full Angle form will interfere with the subsequent text information extraction, so it is changed to the half Angle form. 2) Remove non-Chinese characters such as url links, English Numbers, Chinese and English punctuation marks, blank items, special symbols, etc. 3) The comments with obvious emotional tendency less than 5 after the above treatment were removed, and these data were meaningless to model training and evaluation.

Secondly, jieba is adopted in this paper to perform word segmentation on the cleaned data. At this time, custom basic sentiment lexicon and network sentiment lexicon are added to achieve an ideal effect of word segmentation. For the part of basic dictionary, this paper selects HowNet sentiment lexicon and Dalian Institute of Technology sentiment lexicon. Through artificial judgment and selection, the emotional words that are not commonly used in the review text and whose emotional tendency is unclear are removed, and the positive and negative emotional words with positive meaning (3523) and negative meaning (4869) are merged into them. In the online lexicon, through collecting the popular online lexicon provided by Sogou input method and the emotional words of various social networking sites, sorting and marking the part of speech, a total of 236 words were collected, among which 85 were positive words and 151 were negative words.

Finally, proceed to stop the word operation. This paper selects the stop-word list of Harbin Institute of Technology to remove articles, conjunctions, prepositions and other words without functional information in the text.

2.2. Word2vec—Word Vector Model

In NLP, words are the smallest particle size to form sentences, chapters and documents, so to solve the problem of emotional classification of text, we should start with words. Since words are abstract summaries of human beings, we need to convert characters and symbols into numerical forms and embed them in the mathematical space. This embedding method is word embedding. Word2vec is one of the embedding methods, which can reduce the dimension of input data. Word2vec includes two models: CBOW model and Skip-gram model. The model is divided into input layer, hidden layer and output layer. Skip-gram model is selected in this paper, which uses the selected current word to predict the k words around. In this paper, $k = 5$. The principle is to use the conditional probability value of the intermediate word vector to solve the probability value of the upper and lower window words.

The essence of Word2vec is a neural network language model, based on which distributed word vector training is carried out. It requires a large amount of corpus training to find the relationship between words. After we have trained a word vector model, new corpus will flow into the database, and incremental training will come in handy. Therefore, we increase the scale of the corpus of training word vectors, that is, add SogouCA [10], so that the model can better understand the semantic relationship between words. We can calculate the similarity of two words, find the word with the highest similarity, and find the least relevant outliers in some words.

2.3. Word2vec—The Incremental Training

In deep learning, the effect of the model depends on the size of the training data set, sample diversity and algorithm design. The larger the data set is, the more

suitable it is for deep model learning to improve its generalization ability. In this experiment, the collected 60,000 consumption comments were first used as training data to obtain the word vector model, then the corpus scale was increased, and Sogou whole network news data (SogouCA) was used for incremental training. The SogouCA corpus includes news information from 18 channels including international, domestic, social, sports and entertainment, and is provided in the form of URL and body information. Sogou laboratory provided researchers with data sets of different versions and sizes for downloading. In this experiment, the complete version was downloaded and the news text data with the size of 1.84G was extracted and read. Process for:

Input: call the update_model(), load the old model, update the vocabulary and call the training function for the new corpus after preprocessing

Output: Get_model () gets the Word2vec model and saves it

Step 1: Convert the corpus format from GBK to UTF-8 encoding, obtain the content, and save it in news.txt.

Step 2: Clean off excess tags, load the custom dictionary Jieba participle, and stop using words

Step 3: Read the file contents, use the split_file () function to make the document slices, divide into several documents with 10,000 as a partition, and save the split contents

Step 4: Call Continue_train () to change the corpus into a set of sentences, and load the old model on the updated corpus for training and saving the model

---

3. The Combinatorial Model CNN-BiLSTM-Attention

Firstly, CNN is used to extract phrase features. The CNN model mainly includes convolutional layer, pooling layer, full connection layer and output layer. The main processes are:

1) Input layer: Receives the embedding layer's text matrix as input. Suppose there are n words in a sentence and the vector dimension is d, $d = 150$ in this paper. Is a vector of words for each word in the sentence. Is the input word vector matrix, expressed as:

$$x_{1:n} = x_1 \oplus x_2 \oplus x_3 \oplus \cdots \oplus x_n \qquad (3.1)$$

2) Convolutional layer: Feature extraction is completed through one-dimensional convolution operation. In this paper, three different kernels with sliding Windows of [3] [4] [5] are adopted to set 512 filters for each filter operation, and C_i is used to represent the extracted features. Represented by:

$$C_i = f\left(\omega \cdot x_{i:i+s-1} + b\right) \qquad (3.2)$$

where, is the convolution kernel, S is the size of the convolution kernel, $x_{i:i+S-1}$ is the word vector consisting of i to $i+S-1$, b is the bias term, and the characteristic C_i is obtained through the nonlinear function. When the convolution is complete, you get the eigenmatrix C. Represented by:

$$C = \left[C_1, C_2, \cdots, C_{n-s+1}\right] \qquad (3.3)$$

3) Pooling layer: Pool the obtained feature matrix C to get the local maximum value. MaxPooling method is adopted in this paper. As follows:

$$P = \max\left(C_1, C_2, \cdots C_{n-s+1}\right) = \max\left\{C\right\} \tag{3.4}$$

4) Full connection layer: Since the input of BiLSTM network must be serialized data, it is necessary to connect the P_i after pooling into vector Z. Represented by:

$$Z = \left\{P_1, P_2, \cdots, P_n\right\} \tag{3.5}$$

Use Z as input to the BiLSTM network.

Secondly, the matrix Z is taken as the input of BiLSTM. BiLSTM is a neural network composed of positive and negative output state connection layers of forward and backward LSTM. After BiLSTM feature extraction, the relationship between contexts can be fully learned for semantic coding. BiLSTM is designed to solve the problems of long-term dependence and gradient disappearance. The LSTM differs from a simple RNN in that it does not have a single tanh layer inside, but four interacting layers. The LSTM realizes the selective passing of information through the structure of the gate, mainly through a neural layer of sigmoid and a point-by-point multiplication operation. Each LSTM has a forget gate, an update gate and a output gate to implement protection and control information.

Starting at a certain point, input $x_0, x_1, x_2, \cdots, x_i$ in sequence. The forward propagation layer reads from x_0 to x_i in turn and obtains the forward vector, while the back propagation layer reads from x_i to x_0 in turn and obtains the reverse vector. At time t, the forward hidden state and the reverse hidden state are respectively expressed as (3.6) and (3.8):

$$\overrightarrow{S_t} = \text{LSTM}\left(\overrightarrow{S_{t-1}}, Z_t\right) \tag{3.6}$$

$$\overleftarrow{S_t} = \text{LSTM}\left(\overleftarrow{S_{t-1}}, Z_t\right) \tag{3.7}$$

$$S_t = W_t \overrightarrow{S_t} + V_t \overleftarrow{S_{t-1}} + b_t \tag{3.8}$$

where, $\overrightarrow{S_t}$ and $\overleftarrow{S_t}$ are the hidden states of forward and backward; LSTM is the LSTM function; Z_t is the input of BiLSTM at time t; W_t and V_t are weight matrix and bias quantities.

Finally, in order to highlight the importance of different words to the emotional classification of text, the Attention mechanism layer is introduced in the CNN-BiLSTM-Attention model to further extract text features and highlight key information of text. Since adjectives with emotional color are crucial to the classification of emotions, a higher probability weight is assigned to this part of word vectors. This article belongs to Feedforward attention mechanism, which expressed as formula (3.9) to (3.11).

$$h_t = \sigma\left(S_t\right) \tag{3.9}$$

S_t Is the eigenvector output by BiLSTM, σ is the attention learning function,

which is determined by S_t here, so is tanh, h_t is the weight of the generated attention calculated.

Attention weight normalization: Softmax function was used for normalization to generate attention probability vector, which can be expressed as:

$$a_t = \frac{\exp(h_t)}{\sum_{i=1}^{m} \exp(h_t)} \tag{3.10}$$

The fusion representation is obtained: multiply the attention probability a_t and the hidden state semantic encoding S_t, and then the weighted sum is used to allocate the attention weight. Finally, the fusion feature Q is obtained. Represented by:

$$Q = \sum_{i=1}^{m} a_t \cdot S_t \tag{3.11}$$

The output of Attention is taken as the input of the full connection layer, and sigmoid is used to judge the sentiment of sentences. The attention model allows each word in the sentence to obtain different weights. In the text sentiment analysis, this operation can strengthen the importance of emotion words again, so as to ensure the accuracy of the results.

4. Experiment

4.1. Data Source

Experiment using a data set consists of two parts, part is used for neural network model of sentiment analysis Chinese consumer reviews data set [11], the data set is by making the user on the e-commerce sites collected, containing more than 60,000 comments data, including positive and negative comment on the article 30,000, including books, 10 categories, such as mobile phone, computer, hotel, etc. In this experiment, after the data were cleaned, the positive and negative comments were randomly divided into 13,000 pieces for the experiment, among which 80% was used as the training set and 20% as the test set. The other part is the SogouCA corpus for incremental training.

4.2 Parameter Setting and Model Evaluation

The training word vector and network model parameters are shown in Table 1.

On the test set, the value of accuracy, precision, recall and F1 is used to evaluate the model. The higher the value is, the better the model classification ability is, as shown in Equations (4.1)-(4.4). For the binary classification problem in this paper, the real value of the test set label is combined with the predicted value of the model, and four conditions can be obtained, as shown in Table 2.

$$\text{accuracy} = (TP + TN)/(TP + TN + FP + FN) \tag{4.1}$$

$$\text{precision} = TP/(TP + FP) \tag{4.2}$$

$$\text{recall} = TP/(TP + FN) \tag{4.3}$$

$$f1 = 2 \times (\text{precision} \times \text{recall})/(\text{precision} + \text{recall}) \tag{4.4}$$

Table 1. Parameter settings.

Neural network model parameters			
parameter	values	parameter	values
Vec_dim	150	Lr	0.001
Vec_window	5	Epoch	10
Cnn_filter*num	[3, 4, 5]*512	Batch_size	64
Cnn_activation	Relu	Loss	binary_crossentropy
LSTM_input	150	Optimizer	Adam
Bilstm_dim	128	Dence	2

Table 2. Four cases of true and predicted values.

Real value \ Predictive value	positive	negative
positive	true positive (TP)	false negative (FN)
negative	false positive (FP)	true negative (TN)

Experiment 1:

Three different feature extraction methods were compared: 1) word segmentation based on sentiment lexicon and Word2vec incremental training (SL-W2V-Plus); 2) word segmentation based on sentiment lexicon and Word2vec non-incremental training method (SL-W2V); 3) Word2vec training word vector method (W2V) was used for analysis instead of loading sentiment lexicon. Feature input models are CNN, LSTM and BiLSTM models, and different accuracy, precision, recall rate and F1 values are obtained. As shown in Table 3.

1) The feature engineering method of Word2vec incremental training word vector (SL-W2C-Plus) combined with emotional dictionary auxiliary segmentation is effective. In the CNN model, the method based on SL-W2C-Plus-CNN shows that the four evaluation indexes are the highest in the test data, among which the accuracy is 0.8912, the accuracy is 0.8957, the recall rate is 0.8839, and the F1 value is 0.8842. In addition, the performance of each index in the proposed method is about 0.2% - 0.8% higher than that of SL-W2C-CNN method, and about 0.7% - 1.7% higher than that of W2C-CNN method. In addition, SL-W2C-CNN method is also 0.4% to 0.8% higher than W2C-CNN method. It shows that incremental training can help the vector model to obtain more words representation, and the use of emotion dictionary auxiliary word segmentation can also guarantee the accuracy of word segmentation, and the use of sentiment lexicon auxiliary word segmentation combined with incremental training of word vector model is the best method.

Table 3. Comparison results of three neural network models with different feature engineering methods.

Models	Accuracy	Precision	Recall	F1-measure
W2V-CNN	0.8747	0.8777	0.8764	0.8754
SL-W2C-CNN	0.8832	0.8761	0.8823	0.8772
SL-W2C-Plus-CNN	**0.8912**	**0.8957**	**0.8839**	**0.8842**
W2V-LSTM	0.8879	0.8670	0.8906	0.8891
SL-W2C-LSTM	**0.8941**	0.8802	0.8917	0.8853
SL-W2C-Plus-LSTM	0.8936	**0.8911**	**0.8961**	**0.8905**
W2V-BiLSTM	0.8897	0.8951	0.8820	0.8789
SL-W2C-BiLSTM	0.8948	0.8940	**0.9010**	0.8875
SL-W2C-Plus-BiLSTM	**0.9034**	**0.8989**	0.8913	**0.8954**

2) Similarly, through observation and calculation results, it is found that the improved feature engineering method is also effective in improving the performance of LSTM and BiLSTM models. In SL-W2C-Plus-LSTM method, F1 value is 0.8905 and recall rate is 0.8961, achieving the highest value, and the performance is improved by 0.5% - 1.0% compared with the other two methods. In SL-W2C-Plus-BiLSTM method, the accuracy value is 0.9034, F1 value is 0.8954, and accuracy value is 0.8989, achieving the highest value, and the performance is improved by 0.3% to 1.7% compared with the other two methods.

Figure 1 and Figure 2 more intuitively show the comparison of three models based on different feature engineering methods under two comprehensive evaluation indexes of accuracy rate and F1 value.

As can be seen from Figure 1, the accuracy of the three models increases successively, among which the increase of CNN's accuracy is the most obvious, indicating that the addition of sentiment lexicon auxiliary word segmentation and incremental training word vector are of great help to CNN in obtaining and learning features. The LSTM and BiLSTM models work equally well with SL-W2C and SL-W2C-Plus, and from another point of view, they are a significant improvement over the traditional Word2vec only method to obtain word vectors.

As can be seen from Figure 2, the F1 value of BiLSTM has the most obvious improvement effect, achieving the maximum value of the three models. And then LSTM and CNN. In addition, method 2 (SL-W2C) and the addition of sentiment lexicon auxiliary participle are also helpful to improve the classification effect of the three models.

Experiment 2:

The experiment set up a number of comparison experiments, including the traditional machine learning Support Vector Machine algorithm and Neural Network comparison, single neural network and hierarchical neural network comparison. The word vector model trained by Word2vec is used for network input. The comparison results are shown in Table 4.
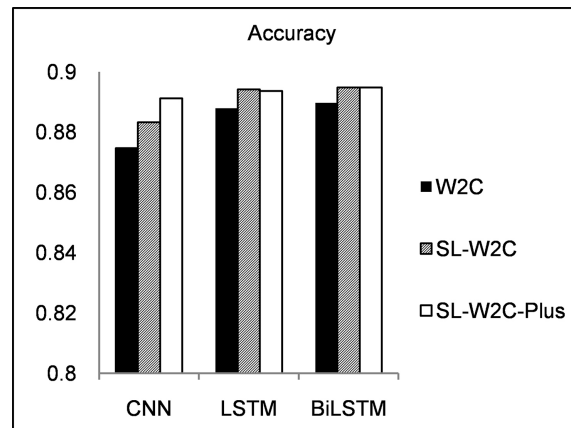
Figure 1. Comparison of the accuracy of engineering methods with different features.
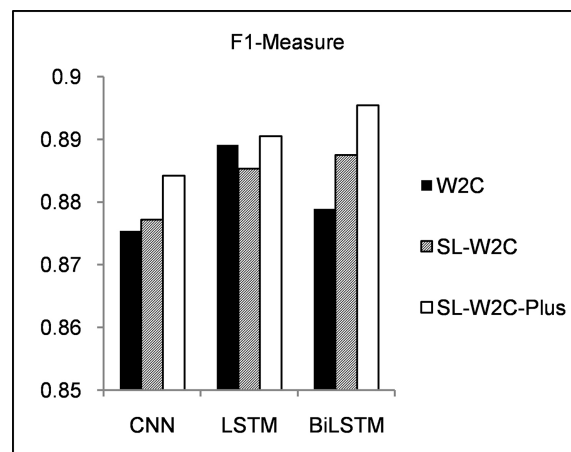
Figure 2. Comparison of F1 values of engineering methods with different features.

Table 4. Models comparison results.

Model	Precision	Recall	F1	Accuracy
SVM	0.8480	0.8527	0.8503	0.8541
CNN	0.8957	0.8839	0.8842	0.8912
LSTM	0.8911	0.8961	0.8905	0.8936
BiLSTM	0.8989	0.8913	0.8954	0.9034
CNN-BiLSTM	0.9106	0.9092	0.9148	0.9051
CNN-BiLSTM-Att	**0.9170**	**0.9203**	**0.9153**	**0.9176**

Table 4 shows the comparison results of six groups of models. In terms of the two comprehensive evaluation indexes F1 and accuracy, the F1 value of CNN-BiLSTM-Attention model reached 91.53% and the accuracy rate reached 91.76%, both of which were better than other models. Although SVM has achieved good classification effect, the other 5 groups of neural network models

are obviously better than SVM. Compared with the models in group 2 and group 3, group 4 and group 5 show the advantages of combined network in feature extraction, because the deep learning of CNN on word vectors is conducive to the reprocessing of CNN feature extraction by BiLSTM. By comparing 5 and 6, it can be seen that adding attention mechanism on the basis of the combination model can effectively improve the accuracy of classification, because attention assigns different weights to features and enables the model to learn that there is a distinction between different features, which is helpful for the model to master important features.

5. Conclusions

In this paper, an algorithm (SL-W2V-Plus) module based on sentiment lexicon word segmentation combined with Word2vec incremental training is proposed for feature selection and calculation of word vectors. Experimental results show that the word vectors obtained by the proposed method are input into the neural network model to obtain the F1 value with higher accuracy. It is proved that this method can obtain higher quality preprocessing results, which is of great significance to the subsequent model feature learning and classification.

In this paper, the combined network CNN-BiLSTM-Attention is selected as the text sentiment classification model, which has higher F1 value and accuracy than other baseline models. Experimental results demonstrate the effectiveness of the proposed method. It provides an idea for text sentiment analysis based on deep learning.

From a business point of view, sentiment-analysis technology makes a big difference. When people want to buy a new product, choose a restaurant to eat, or plan to see a new movie, they can "plant content" by looking at the comments of "people who have been there before". In addition, in recent years the birth of Weibo, Zhihu, Douban and other social media because of its diversified content classification is attracting more and more young people to join, the applications of text, great reviews data hidden behind the huge resources, mining the rich text content, and help the government to collect public opinion on hot issues, correctly guide the emergency or the people's livelihood issues such as the development direction of public opinion. It is also helpful for relevant departments to monitor cyber violence, judge positive and negative comments with the method of sentiment analysis, and strengthen the control over negative comments to reduce the harm caused by improper comments to others.

The deep learning model is still in its infancy in Chinese emotion analysis. Chinese language features such as complex sentence structure, complex system and diverse expression methods, which hinder the development of text emotion analysis to some extent. The quality of feature word vectors has a certain influence on the learning of subsequent network models. The next step is to learn the Bert model pre-trained word vectors proposed by Google. In the selection of the model, the simplified structure network GRU of LSTM should also be tried for

experiment and result analysis.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

[1] China Internet Network Information Center (CNNIC) (2020) 45th Statistical Report on Internet Development in China.

[2] Sonawane, S.L. and Kulkarni, P.V. (2017) Extracting Sentiments from Reviews: A Lexicon-Based Approach. 1st *International Conference on Intelligent Systems and Information Management* (*ICISIM*), Aurangabad, 5-6 October 2017, 38-43. https://doi.org/10.1109/ICISIM.2017.8122144

[3] Taj, S., Shaikh, B.B. and Fatemah Meghji, A. (2019) Sentiment Analysis of News Articles: A Lexicon Based Approach. 2019 2nd *International Conference on Computing, Mathematics and Engineering Technologies* (*iCoMET*), Sukkur, 30-31 January 2019, 1-5. https://doi.org/10.1109/ICOMET.2019.8673428

[4] Wongkar, M. and Angdresey, A. (2019) Sentiment Analysis Using Naive Bayes Algorithm of the Data Crawler: Twitter. 2019 4th *International Conference on Informatics and Computing* (*ICIC*), Semarang, 16-17 October 2019, 1-5. https://doi.org/10.1109/ICIC47613.2019.8985884

[5] Kim, Y. (2014) Convolutional Neural Networks for Sentence Classification. *Proceedings of the* 2014 *Conference on Empirical Methods in Natural Language Processing* (*EMNLP*), Doha, Qatar, October 2014, 1746-1751. https://doi.org/10.3115/v1/D14-1181

[6] Kalchbrenner, N., Grefenstette, E. and Blunsom, P. (2014) A Convolutional Neural Network for Modelling Sentences. *Proc. of the* 52nd *Annual Meeting of the Association for Computational Linguistics*, Baltimore, Maryland, Vol. 1: Long Papers, 655-665. https://doi.org/10.3115/v1/P14-1062

[7] Alayba, A.M., Palade, V., England, M., *et al.* (2018) A Combined CNN and LSTM Model for Arabic Sentiment Analysis. *Lecture Notes in Computer Science*, **11015**, 179-191. https://doi.org/10.1007/978-3-319-99740-7_12

[8] Ji, H., Rong, W., Liu, J., Ouyang, Y. and Xiong, Z. (2019) LSTM Based Semi-Supervised Attention Framework for Sentiment Analysis. 2019 *IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computing, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation* (*SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI*), Leicester, 19-23 August 2019, 1170-1177. https://doi.org/10.1109/SmartWorld-UIC-ATC-SCALCOM-IOP-SCI.2019.00218

[9] Mikolov, T., Sutskever, I., Chen, K., *et al.* (2013) Distributed Representations of Words and Phrases and Their Compositionality. *Proc of the* 26th *International Conference on Neural Information Processing Systems*, Curran Associates Inc., USA, 3111-3119.

[10] Sogou Tech-Oriented News Laboratory Data. http://www.sogou.com/labs/resource/ca.php

[11] Sentiment Analysis in Chinese Corpus. https://github.com/SophonPlus/ChineseNlpCorpus/tree/master/datasets/online_shopping_10_cats