A Sentiment Analysis Method Based on BLSTM and CNN Fusion

Liedong Yin, Xia Ye, Junping Yao
Xi’an Research Institute of High Technology, Xi’an, China
Email: yinld921@qq.com

Abstract. Sentiment analysis is a research hotspot in natural language processing in recent years. This paper proposes a sentiment analysis method that integrates LSTM and CNN in view of the fact that most existing sentiment analysis methods mix semantics and emotions. This method divides the text into semantic space and emotional space, and uses the attention model to train in two dimensions, and fuses the feature representations of the two to construct the final representation of the text. The results show that the model has achieved good results in multiple emotional review data.

1. Introduction

With the rapid development of social media, social networks such as Weibo and WeChat are gradually changing people's lives. More and more people are willing to share their views on events on social networks, not just passively browsing and receiving information[1]. Sentiment analysis of the commentary text helps to understand the public's emotional state, timely access to the public's opinions and attitudes, and has important implications for public opinion control, product marketing, and public opinion surveys.

The existing sentiment analysis techniques are mainly divided into three types: methods based on emotional knowledge, methods based on feature classification, and methods based on deep learning. The method based on emotional knowledge is mainly to construct some emotional dictionaries. Through emotional dictionaries, a certain combination of the emotional words in the text is calculated to realize the sentiment analysis of the text[2,3]. When people express emotions, they often use some special emotional symbols to express their emotions. The commonly used emotional symbols are mainly emotional words with emotional polarity. More often than not, people will use emotional words in conjunction with degree adverbs or negative words to enhance or diminish emotions. In addition to the basic emotional words, each platform's emotional expression includes its own facial expressions and some popular online language. In this paper, emotional words, degree adverbs, negative words and network terms are collectively referred to as emotional symbols, which are used as an important part of evaluating text sentiment.

The feature-based classification method is mainly used text analysis technology to extract the relevant language features contained in the text, and used machine learning methods to treat the sentiment analysis as a classification problem[4,5], but the method relies on the emotional feature extraction method. Larger sex, the accuracy of the extraction of emotional characteristics will directly affect the final classification results.

The method based on deep learning mainly uses the word vector representation technology to express the words in the text, and then constructs the semantic representation of sentences and texts.
Based on the semantic representation of sentences and texts, the deep learning model is used to learn the emotions contained in the texts, so as to realize the analysis of text emotions. The deep learning models commonly used in sentiment analysis include: convolutional neural network (CNN)\(^6\), recurrent neural network (RNN)\(^7\), long short-term memory (LSTM)\(^8\), etc.

Most of the existing sentiment analysis methods are based on deep learning methods. Texts are coded as a whole. The effect of emotion symbols is not enough, and the importance of emotional symbols is not highlighted. The traditional method based on sentiment lexicon is Over-reliance on the role of the emotional dictionary does not take into account the overall semantic relationship of the text. In order to solve the above problems, I introduce the attention model\(^9-10\) that expresses excellent performance in the machine translation task to the sentiment analysis task, extracting the text into two spaces, i.e. semantic space and emotional space.

2. Related work

2.1. Attention Model

AM initially achieved very good results in the field of image processing. So Bahdanau et al.\(^9\) began to study the introduction of AM models into the NLP field. They believed that the fixed-length intermediate semantic vector generated by the Decoder in the traditional Encoder-Decoder model is Improve the bottleneck of neural machine translation model performance. Therefore, Bahdanau et al. proposed an alignment model to calculate the alignment probabilities of the output state at the previous moment and the input state at each moment, and applied it to the translation system of English and French, and achieved good results. Subsequently, the attention-based Encoder-Decoder model has been applied in many fields of natural language processing, such as short text dialogue, text summary generation, text classification, and syntactic analysis.

With the wide application of attention models in the field of natural language processing, various improved models have emerged. For example, Luong et al.\(^10\) proposed a local attention model for the attention-aligned global alignment model. The approximate position of each, and then extend the length of the window to the left and right, calculate the alignment probability of each word within the specified window range. Hermann et al.\(^11\) according to people's reading comprehension topic, often read the complete question and then go to the document to find the answer rather than the actual situation to go to the document to find the answer after reading a word, will be modified in an alignment manner The pattern for aligning the entire question with the words in the document.

2.2. Deep Learning Network

Deep learning has gradually become the hottest branch of machine learning. It can obtain representations of different levels of abstraction of the raw data through multiple levels of learning, and can automatically learn features, thereby improving the accuracy of classification or prediction tasks. The convolutional neural network (CNN) proposed by LeCun et al.\(^12\) has achieved very good results in image tasks. Recently, the CNN model has been proved to have good results in many NLP tasks, such as part-of-speech tagging and sentence construction. Module and sentence classification and so on. The RNN model has been widely applied to the NLP task as another branch of the deep learning model. The advantage of the RNN is that it can store historical information in a series of hidden neurons.

There have been some studies on fused CNN and RNN models, for example, the method of circular convolutional neural network proposed by Pinheiro et al.\(^13\) for image representation and entity recognition. This combination method also has excellent results in natural language processing tasks. Kalchbrenner et al.\(^14\) achieved good results in the dialogue behavior recognition task by incorporating the convolutional structure and the RNN model without any support from the feature engineering.
2.3. BLSTM model
The BLSTM model is a combination of the forward LSTM[15] model and the reversed LSTM model, capable of capturing both positive and negative semantic information at the same time. The LSTM controls the flow of information in the current network through a memory unit $c$ and three gate structures (input gate $i$, forgotten gate $f$, and output gate $o$). Specifically, the information content $c_t$ of the current time memory unit and the output information amount $o_t$ are the input $x_t$ at the current time, the state $f_t$ of the gate at the current moment, the information content $c_{t-1}$ of the memory unit at the previous moment, and the output $o_{t-1}$ at the previous moment. In BLSTM, the relationship between the various structures can be simply expressed as Formula (1) (2):

$$
\tilde{c}_t, \tilde{o}_t = G(\tilde{c}_{t-1}, \tilde{o}_{t-1}, \tilde{f}_t, x_t)
$$

(1)

$$
\tilde{c}_t, \tilde{o}_t = G(\tilde{c}_{t-1}, \tilde{o}_{t-1}, \tilde{f}_{t-1}, x_t)
$$

(2)

Among them, $G$ represents the relationship between the structures in the LSTM model. In the BLSTM, the output state at a time $t$ is composed of the forward LSTM of the current time and the output of the reverse LSTM and can be expressed as $o_t = [\tilde{c}_t, \tilde{c}_t]$.

3. The Sentiment Analysis Model Combining BLSTM and CNN

3.1. Emotional Dictionary Construction

3.1.1. Emotion Basic Dictionary Construction
The above constitutes an emotional basic dictionary. In this paper, through the collation of existing emotional knowledge bases, the emotion words in the "Feeling Ontology Library" of Dalian University of Technology and "HowNet" constructed by Zhendong Dong are merged; degree adverbs are based on the degree adverbs list in "HowNet". Construction is carried out; negative words and network terms are constructed on the basis of negation words constructed by Song et al. [16] and network term lists (NetLex). The above constitutes an emotional basic dictionary.

3.1.2. Emotion extension dictionary construction
Emotional basic dictionary is used to identify emotion words in the comment text, and an emotion expansion dictionary is automatically recognized based on this. For the extraction of emotional symbols, this paper has formulated the following extraction rules:

Rule 1: If the current word is an adverb of degree, and the next word of the current word is an emotional word, then the current adverb and the emotional word are added as a whole to the extended dictionary, and if they exist, they are ignored.

Rule 2: If the current word is a negative word, and the next word of the current word is an emotional word, the current negative word and emotional word are added as a whole to the emotional expansion dictionary, if they already exist, they are ignored; or the current word is underneath. A term is a degree adverb, and the next word of the degree adverb is an emotional word. Then, the current negative word, degree adverb, and emotional word are added into the extended dictionary as a whole. If they exist, they are ignored.

3.2. The Sentiment Analysis Model Combining BLSTM and CNN
In this paper, BLSTM and CNN are used to train corpus, and attention mechanism is introduced to guide the fusion of word vectors. At the higher level, the vector is merged with the generated emotion space. Finally, it is handed to classifier training. The overall architecture of the model is shown in Figure 1.

3.2.1. Word vector layer
The word vector layer is composed of semantic space and emotional space, in which the words of emotional space come from the extended emotional dictionary. The acquisition of words and their emotional symbolic words can be viewed as a dictionary search. Dictionary $R^{d \times m}$ is a CBOW word vector model proposed by
Google for learning. Where \( d \) represents the dimension of the word vector and \( N \) represents the number of words in the dictionary. For a text sequence \( T = \{W_1, W_2, \ldots, W_n\} \), the word vectors of the words in the text are concatenated, and the word vector representation of the entire text sequence can be obtained. The splicing manner is as shown in formula (3).

\[
R_S = v_1 \oplus v_2 \oplus \cdots \oplus v_n
\]  

(3)

Among them, \( v_i \in \mathbb{R}^{d \times m} \) indicates that \( W_i \) corresponds to the value in the dictionary, and \( \oplus \) indicates the row vector splicing operation. Words of words in the emotional space are also stitched using the formula (3). The dimensions of \( R_S \) and \( R_E \) are the number of words in the text and the number of words in the emotional space.

3.2.2. Semantic acquisition layer
The LSTM type RNN network has a natural advantage in handling sequence tasks such as sentences or texts. It can store and use the historical information of the sentence, so LSTM has a strong long-range semantic capture capability, so as to establish the semantic relationship between clauses or different words. However, because the dataset is small or the short sentences are too large, LSTM cannot extract sentence-level features very well. Therefore, the results in some comparative experiments are only better than ordinary RNN models and even less accurate than the network models such as CNN. A layer of CNN network model can handle the information in the window very well. For the information outside the window, only by increasing the number of layers to solve the semantic association, but this also increases the complexity of the network structure, but also to the parameter tuning with New burden. In fact, the Weibo text commentary texts are of short or long length, and therefore the integration of LSTM and CNN at this level makes them complementary. Emotional spatial data is small and small, so CNN training is used.
The method of attention mechanism fusion is as Equation (4):
\[ C_a = W_C S_C + W_B L_C \] (4)

Wherein \( W_C \) and \( W_B \) respectively weight the CNN and LSTM weight matrix, \( S_C \) the CNN output, and \( L_C \) the LSTM output.

### 3.2.3. Semantic synthesis layer

The main task of the synthesis layer is to combine the \( S_E \) and \( S_S \) obtained from the \( R_E \) and \( R_S \) to construct the semantic representation vector of the text as a whole. In this paper, in order to simplify the calculation of the model, only SE and SS are combined in a row-by-line manner to form a \((r_S + r_E) \times c\) matrix \( S \), and the text is semantically represented. Among them, \( r_S \) and \( r_E \) represent the number of rows of \( S_S \) and \( S_E \), and \( c \) represents the number of columns of \( T \) and \( E \).

### 3.2.4. Affective computing layer

The main task of the emotional computing layer is to construct a sentiment classifier, obtain the semantic representation of the Weibo text, and score the score vector for each emotional tag, and output the final emotional tag of the text. This paper uses Softmax classifier to construct the score vector of each emotion tag, and converts it into a conditional probability distribution.

\[ P_i(x) = \frac{\exp(x_i)}{\sum_{j=1}^{C} \exp(x_j)}, \quad i = 1, 2, \ldots, C \] (5)

Among them, \( C \) represents the number of emotional tags. In order to better train the model, we use the cross-entropy loss function to measure the gap between the true probability distribution \( P_i(d) \) of the emotion tag and the predicted probability distribution \( P_i(d) \).

\[ L = -\sum_{d \in T} \sum_{i=1}^{C} P_i(d) \log(P_i(d)) \] (6)

Among them, \( T \) is the training data set, \( d \) is a text in \( T \). This article uses the cross-entropy loss function of equation (6) to use the back propagation mechanism to train and update parameters in the model.

### 4. Experiments

#### 4.1. Data set

This experiment introduces the three types of corpus in the corpus of Chinese emotions collected by Professor Songbo Tan. Each type of data contains 2,000 positive and negative texts. In order to test the advantages of this model for long and short sentences, the experimental data is partially deleted. The deleted strategy is the length of the current text. You need to get the short text and delete the objective text. The data obtained is shown in Table 1.

|                  | Hotel | phone | Book |
|------------------|-------|-------|------|
| Average length   | 167   | 108   | 254  |
| Standard deviation| 20    | 13    | 153  |
| Total number     | 1387  | 1231  | 1537  |

#### 4.2. Analysis of model results

In this experiment, the sigmoid output activation function is used on the binary dataset. The corresponding binary_cross_entropy (binary cross-entropy loss function) is used during model compilation. When training the data set, the softmax activation function is used at the output layer, and the corresponding loss function is a category cross-entropy function. Such loss function and activation function one-to-one correspondence mechanism can make the error of each output neuron exactly equal to the error between its output and the marked real data when the error propagates.
To further avoid training overfitting, the dropout mechanism was added to the input and output layers of the word vector. After the comparison, the dropout ratio was set to 0.2, and five-fold cross validation was used during the experiment.

| Table 2. Experimental result |
|-----------------------------|
|                | Hotel | phone | Book |
| LSTM           | 85.6  | 84.5  | 85.1 |
| CNN            | 85.5  | 84.7  | 84.2 |
| BLSTM          | 85.8  | 84.7  | 84.9 |
| E-BLC          | 86.8  | 85.8  | 85.9 |
| ES-BLC         | 87.3  | 86.4  | 87.5 |

It can be seen from Table 2 that LSTM and CNN have a slight advantage over LSTM when the length of the text is not large, but in the case of a large length, the advantage of LSTM is even more pronounced. Compared with LSTM, BLSTM gains reverse semantics and therefore performs better. E-BLC removes the emotional space in the ES-BLC model and preserves only the semantic space. Therefore, the model does not have a synthetic layer. By comparison, E-BLC performed better than LSTM, CNN, and BLSTM on all three datasets, indicating that the model can complement the advantages of BLSTM and CNN and improve the classification effect. The ES-BLC model added to the emotional space has a greater improvement than the E-BLC model, which shows that the part of adding emotional space does strengthen the emotional information and the obtained classification effect is better.

5. Conclusion

This paper proposes an ES-BLC model and uses existing sentiment analysis resources to construct an emotionally extended dictionary containing emotional words, degree adverbs, negative words, emoticons, and commonly used network terms, and uses the dictionary to extract emotional words in texts. Form emotional space. By giving the semantic space to the BLSTM and CNN and through the attention model fusion, the emotional space is trained on the CNN and finally combined to form a vector representation of the text, which effectively enhances the model's ability to capture the text's emotional semantics, thereby improving the text Sentiment analysis performance.

References

[1] Ran Li,Zheng Lin,Hailun Lin,Weiping Wang and Dan Meng. Text Sentiment Analysis Review[j].Journal of Computer Research and Development,2018,55(01):30-52.
[2] Taboada M, Brooke J, Tofiloski M, et al. Lexicon-based methods for sentiment analysis[J]. Computational Linguistics, 2011, 37(2):267-307.
[3] Wiebe J, Wilson T, Cardie C. Annotating Expressions of Opinions and Emotions in Language[J]. Language Resources & Evaluation, 2005, 39(2/3):165-210.
[4] Pang, Bo, Lee, Lillian. Opinion Mining and Sentiment Analysis[J]. Foundations & Trends in Information Retrieval, 2008, 2(1–2):1-135.
[5] Pang B, Lee L, Vaithyanathan S. Thumbs up?: sentiment classification using machine learning techniques[J]. Proceedings of Emnlp, 2002:79-86.
[6] Kim Y. Convolutional Neural Networks for Sentence Classification[J]. Eprint Arxiv, 2014.
[7] ZHANG Yangsen,JIANG Yuru,TONG Yixuan.Study of Sentiment Classification for Chinese Microblog Based on Recurrent Neural Network[J].Chinese Journal of Electronics,2016,25(04):601-607.
[8] Wang J, Yu L C, Lai K R, et al. Dimensional Sentiment Analysis Using a Regional CNN-LSTM Model[C]/ Meeting of the Association for Computational Linguistics. 2016:225-230.
[9] BAHDANAU D, CHO K, BENGIO Y. Neural machine translation by jointly learning to align and translate[C]/Proceedings of the International Conference on Learning Representations. San Diego, USA, 2015.
[10] Luong M T, Pham H, Manning C D. Effective Approaches to Attention-based Neural Machine
Translation[J]. Computer Science, 2015.

[11] HERMANN K M, TOMÀS K, GREFENSTETTE E, et al. Teaching machines to read and comprehend[C]//Proceedings of the International Conference on Neural Information Processing Systems. Montreal, Canada, 2015:1693-1701.

[12] Le Cun Y, Bottou L, Bengio Y, et al. Gradient-based learning applied to document recognition [J]. Proceedings of the IEEE, 1998, 86 (11): 2278–2324.

[13] Pinheiro P H O, Collobert R. Recurrent convolutional neural networks for scene parsing [J]. 2013: arXiv: 1306. 2795.

[14] Kalchbrenner N, Blunsom P. Recurrent convolutional neural networks for discourse compositionality [J]. Computer Science, 2013.

[15] HOCHREITER S, SCHMIDHUBER J. Long short-term memory[J]. Neural Computation, 2012, 9(8):1735-1780.

[16] SONG K, FENG S, GAO W, et al. Build emotion lexicon from microblogs by combining effects of seed words and emoticons in a heterogeneous graph[C]//Proceedings of the ACM Conference on Hypertext and Social Media. New York, USA, 2015:283-292.