

Research on Sentiment Analysis of Micro-blog based on Attention-BiLSTM

Fangxu Yan, Jianhui Wang

School of Mathematics and Systems Science, Shenyang Normal University, Shenyang Liaoning, 110034, China

Abstract: The rapid development of the Internet gives birth to massive data. These data have the characteristics of multiple types, low density, and fast transmission speed. As one of the products of the era of big data, micro-blog provides users with a highly interactive platform. In order to explore the value and sentiment contained in micro-blog comment texts, and solve the problem of semantic loss and over-reliance on manual work in traditional text sentiment analysis, this paper proposes a deep learning model based on Attention-BiLSTM. In the embedding layer, the CBOW method of Word2vec is used to transform the text data into word vectors, and the attention mechanism is integrated to dynamically weight the input word vectors. Secondly, BiLSTM is used to extract text features, then fuse them in concat layer. After that the classification is realized by Softmax whose results of the output text are positive or negative. Finally, by comparing with TextCNN, LSTM and BiLSTM, the result shows that the Attention-BiLSTM model does have better classification effect, strong generalization and practicability.

Keywords: Sentiment Analysis; BiLSTM; Attention Mechanism.

1. Introduction

According to the latest 2023 financial report released by Micro-blog, at the end of the third quarter, the number of monthly active users of Micro-blog reached 605 million, and daily active users reached 260 million. Micro-blog provides a simple way for users to publish content openly and in real time, which makes the interaction among users closely linked. It has become one of the main channels for the public to obtain information, understand current affairs and express their views [1]. Such massive comment data often contains a strong tendency and subjectivity. Therefore, it is of great significance to conduct sentiment analysis on hot search comments in social media.

At present, there are three mainstream methods to realize the sentiment analysis of text. Sentiment dictionary calculates the emotional score based on some specific rules and determine the polarity. Miao proposed the concept of four-tuple extraction, so as to realize the analysis of feature level [2]. Narayanan proposed a classification method based on clauses, result sentences and the whole sentence, and the effect is remarkable [3]. machine learning uses sample data to train the model in order to complete the prediction. Pang proposed a machine learning method to classify the emotional polarity, which is divided into positive and negative [4]. On this basis, Whitelaw uses the vector space model to represent the text, which further improves the accuracy [5]. Deep learning completes the sentiment analysis of the text through a hybrid neural network and a pre-trained model. In 1990, Elman proposed RNN recurrent neural network for processing sequence data [6]. Vaswani proposed a Transformer architecture based entirely on the multi-headed attention mechanism, and for the first time proposed the attention mechanism [7].

According to the experience above, this paper proposes a sentiment analysis model of Micro-blog based on Attention-BiLSTM. The experimental results show that compared with the above similar text experiments, this model does improve the accuracy of text emotion recognition.

2. Model Methodology

The overall structure of the model is shown in Fig.1. The model is mainly consist of Embedding layer, Attention layer, BiLSTM, and Softmax layer. Each level is described in detail below.

2.1. Embedding

In embedding layer, we mainly use Word2vec to process comment data, which could be transformed into vectors in order to raise the dimension. CBOW (continuous bag of words), one of the methods in Word2vec, is used to arrange the context words corresponding to the central word in order, then project the context words in the selected window to the same position, and average the vector. The output result is the word vector of the central word, which is represented as (x_1, x_2, \dots, x_n) and input to the attention layer.

2.2. Attention Layer

The core of the attention layer is the attention mechanism, which performs selective key information processing by imitating the working principle of the human brain. Furthermore, it can give higher weights to words with greater importance. The input vector (x_1, x_2, \dots, x_n) corresponds to (Q, K, V) . Firstly, calculate the similarity for each input vector by multiplying Q and K^T . Then, the similarity is converted numerically by Softmax function, which means to normalize the similarity and reduce the amount of calculation to a certain content [8]. After normalization, multiply the similarity score by V . Finally, the obtained result can be calculated according to Equation (2) in which d_k represents the dimension of vector K .

$$\begin{aligned} & \text{Attention}(Q, K, V) \\ &= \text{Softmax}\left(\frac{Q \cdot k^T}{\sqrt{d_k}}\right)V \end{aligned} \quad (1)$$

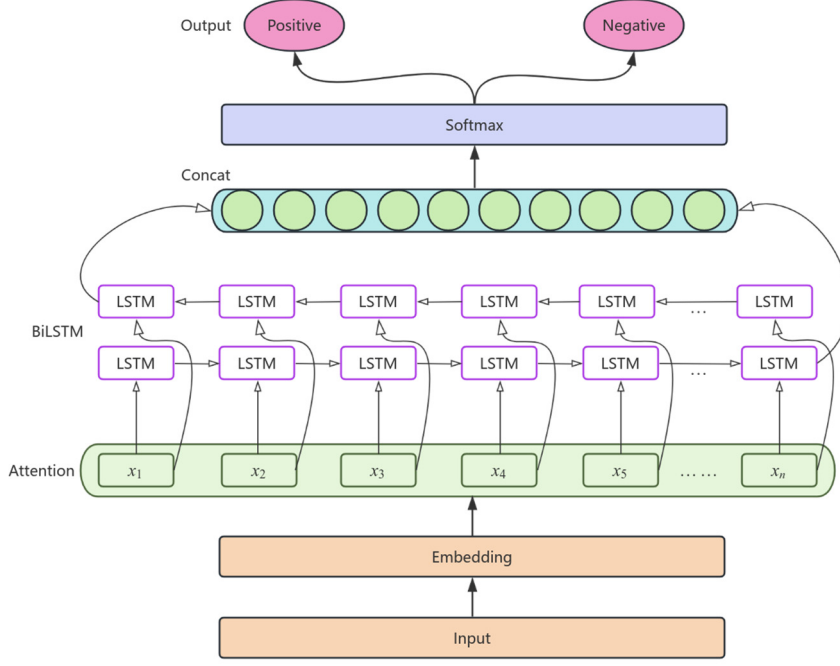


Fig 1. Overall structure of the model based on Attention-BiLSTM

2.3. BiLSTM

The traditional LSTM model can store and use all the previous position states to better describe the dependency relationship between long and short distances, which also avoid the problem of gradient explosion or gradient disappearance [9]. Based on this, BiLSTM introduces forward propagation layer and backward propagation layer as shown in Fig.2. x_t represents input, $h_t^{(1)}$ represents the hidden state of forward propagation, and $h_t^{(2)}$ represents the hidden state of backward propagation layer, denoted in Equation (2) and Equation (3).

$$h_t^{(1)} = \tanh(U^{(1)} \cdot h_{t-1}^{(1)} + W^{(1)} \cdot x_t + b^{(1)}) \quad (2)$$

$$h_t^{(2)} = \tanh(U^{(2)} \cdot h_{t-1}^{(2)} + W^{(2)} \cdot x_t + b^{(2)}) \quad (3)$$

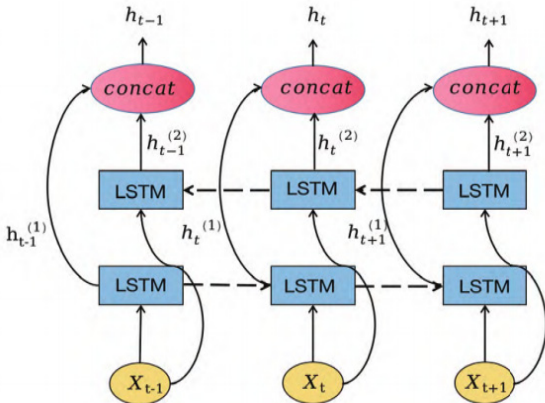


Fig 2. Propagation mechanism of BiLSTM

The output state of two LSTM units is spliced to form a new state vector h_t in Equation (4).

$$h_t = \text{concat}(h_t^{(1)}, h_t^{(2)}) \quad (4)$$

2.4. Softmax Layer

Finally, the Softmax function is used to normalize the generated text vector, and the probability distribution of the text vector under positive or negative types is obtained. In Equation (5), $i = 1$ represents Positive and $i = 0$ represents Negative. The obtained result can be calculated as follow:

$$p_i = \text{softmax}(z_0, z_1) = \frac{\exp(\omega z_i + b)}{\sum_i \omega z_i + b} \quad (5)$$

3. Analysis Experiment

3.1. Datasets

Data sources in this paper come from Github, with a total of 8500 labeled comment data. Among them, there are about 5500 positive comments and 3000 negative comments. Due to the complexity of the composition of the comment text, including Chinese, English, slang, digital symbols, special expressions, vacancy values, redundant values, etc., it is necessary to process the text data.

3.2. Data Preprocessing

First, data is simply cleaned to remove noise such as null values and redundant values. Then jieba is used to divide each data into a set of independent words. This paper integrates the HIT stop vocabulary and Baidu stop vocabulary to form a micro-blog comment stop vocabulary, filters the useless characters and auxiliary words in the data, and inputs the Word2vec model for word vector training.

3.3. Evaluation Index

We introduce Accuracy, Precision, Recall, and F1-Score as

the evaluation indicators of the model as shown in Equation (6) to (9). Among them, TP represents predicting positive text correctly as positive, FN represents the positive text is predicted to be negative, FP represents the negative text is predicted to be positive, and TN represents the negative text is predicted to be negative correctly.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (6)$$

$$\text{Precision} = \frac{TP}{TP + FP} \quad (7)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (8)$$

$$\text{F1 - Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (9)$$

3.4. Parameter Settings

The training set and test set are divided into 80% and 20%. The parameter configuration of the Attention-BiLSTM model is shown as Table 1.

Table 1. Parameter settings of the Attention-BiLSTM model

Parameter	Value
Embedding_size	228
Epochs	15
Max_len	48
Batch_size	500
Learning Rate	0.001

3.5. Contrast Experiment

Here we introduce three sets of contrast experiments, namely LSTM, BiLSTM, and TextCNN, then compare the results with the model proposed in this paper as shown in Table 2. By using multiple convolution kernels of different sizes, the TextCNN model can extract the interrelationships within and between sentences. But the pooling operations in TextCNN will lose the order and location information between the dialogue texts, which is not conducive to the sentiment recognition of the comment text. LSTM model performs better in learning the long-term dependency of sequence data whose accuracy reaches 79.57%. Based on LSTM, BiLSTM has a two-way propagation mechanism, which can better characterize the sequence data in the process of learning, and has better generalization ability. However, the single model neural network is relatively simple, which fails to fully consider the text features. Through comparison, it can be seen that Attention-BiLSTM model has improved the indicators compared with Text-CNN, LSTM, and BiLSTM models. It is further proved that the structure of the composite model is superior to the single model in text sentiment classification, with higher accuracy and stronger universality.

Table 2. Comparison of different models

Model	Accuracy	Precision	Recall	F1-Score
Text-CNN	0.7347	0.6949	0.6564	0.6771
LSTM	0.7957	0.8435	0.8027	0.8225
BiLSTM	0.8217	0.7988	0.8139	0.8198
Attention-BiLSTM	0.9235	0.8950	0.8234	0.8667

4. Conclusion

To analyze the Micro-blog text emotion, we propose the compound model Attention-BiLSTM. The result proves that the model gains much better accuracy, flexibility and feasibility in the field of text sentiment classification. On next stage, the goal is to develop higher indicators and achieve multi-modal sentiment analysis on the existing basis.

Acknowledgments

The authors gratefully acknowledge the financial support from Liaoning Provincial Department of Education Scientific Research Funding Project funds (LFW202004).

References

- [1] Xiong W, Zuo Y, Zhang M, et al. Research on Sentiment Analysis of E-commerce Live Comments based on Text Mining[J]. Frontiers in Computing and Intelligent Systems, 2023, 6(3): 34-36.
- [2] Miao Q, Li Q, Dai R. AMAZING: A sentiment mining and retrieval system[J]. Expert Systems with Applications, 2009, 36(3): 7192-7198.
- [3] Narayanan R, Liu B, Choudhary A. Sentiment analysis of conditional sentences[C]//Proceedings of the 2009 conference on empirical methods in natural language processing. 2009: 180-189.
- [4] Bo Pang, Lillian Lee, Shivakumar Vaithyanathan. Thumbs up? sentiment classification using machine learning techniques [C]// Proceedings of the ACL-02 conference on Empirical methods in natural language processing. 2002: 79-86.
- [5] Whitelaw C, Garg N, Argamon S. Using appraisal groups for sentiment analysis[C]//Proceedings of the 14th ACM international conference on Information and knowledge management. 2005: 625-631.
- [6] Elman J L. Finding structure in time[J]. Cognitive science, 1990, 14(2): 179-211.
- [7] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need [J]. Advances in neural information processing systems, 2017, 30.
- [8] Zhang J, Zhang X, Liu Z, et al. A Network Intrusion Detection Model Based on BiLSTM with Multi-Head Attention Mechanism[J]. Electronics, 2023, 12(19): 4170.
- [9] Zhang B. COVID-19 forecast and bank credit decision model based on BiLSTM-attention network[J]. International Journal of Computational Intelligence Systems, 2023, 16(1): 159.