Sentiment Analysis of E-commerce Comments Based on BERT

Xiaohong Cai[1, a], Hui Cao[1, b], Jingang Ma[1, *]

[1]School of Intelligence and Information Engineering Shandong University of traditional Chinese Medicine, Jinan, 250355, China.

[a]caixhsdutcm@163.com, [b]caohui63@163.com, *ma_jingang@163.com

Abstract

Sentiment analysis is one of the hot research contents of natural language processing, and e-commerce comment sentiment analysis has become a research hotspot. Sentiment analysis is one of the hot research contents of natural language processing. Sentiment analysis of e-commerce comments has become a research hotspot. The current word vector representation model does not contain the context of text. Therefore, this paper proposes a pre training language model SA-BERT based on transformer bidirectional encoder representation. First, the word vector is encoded by Bert to represent the semantic information of the context of the comment text. Then, the attention mechanism is used to extract the text features in a deeper level, understand the semantics of the text information, and complete the sentiment analysis task of e-commerce comment. Finally, experiments on JD mobile comment data set show that SA-BERT model has a good effect in this field.

Keywords

Deep Learning, BERT, Comments, Sentiment Analysis.

1. Introduction

With the development of science and technology, big data is getting more and more attention. People can express their opinions through various channels, and a large amount of textual information such as comments, articles, and messages continue to accumulate on the Internet, forming massive data. The comment data on the Internet contains huge commercial value. Analyzing Internet text data can not only help individuals and businesses, but even help governments make decisions. For individuals, when they need to buy a product, they will find a lot of information about the product's reviews and discussions on the Internet, and then decide whether to buy; for companies, collect evaluation information directly from the Internet and analyze , Not only helps to establish and maintain the corporate image, but also helps to understand the attributes of products or services that are widely concerned by users for targeted improvement, and then to improve the overall efficiency.

Although the huge amount of text data appearing on the Internet has huge commercial value, how to better excavate these potential commercial values is a problem that we need to solve urgently. Taking online shopping as an example, the volume of review data is large, and the effective information in the review data is randomly distributed, which makes it difficult for users to quickly locate the information they care about. Simply browse the online reviews by individuals to obtain a comprehensive evaluation of the product Obviously it is unrealistic. Traditional sentiment analysis only analyzes the overall sentiment tendency of a review. This traditional sentiment analysis not only leads to insufficient information extraction, but also fails to accurately identify the sentiment tendencies of the product that the reviewer cares about.

In recent years, with the development of deep learning and the use of natural language, more and more scholars have adopted deep learning techniques to process massive amounts of text information. Deep learning can automatically learn the essential characteristics of the data through multi-level learning, and can obtain a higher accuracy in the prediction and classification tasks. For example, the commonly used networks are: Recurrent neural network [1], Long and Short-term Memory Network[2], etc.

The main work of this paper is as follows:

1) Use BERT to vectorize the sentence of the comment information, fully considering the impact of each word in the sentence on the words in other contexts and the different meanings of the same word in different contexts.

2) First use the BERT network structure to process the extraction of contextual features in the text, and finally introduce an attention model to assign weights to the extracted information to highlight key information, thereby improving the accuracy of text sentiment analysis.

2. Related Works

2.1 BERT

The generation of BERT has brought a great improvement in the field of natural language processing. The previous models were combined from left to right or from left to right and right to left training, and BERT (shown in Fig.1) used a multi-layer Transformer model[3], the information of each word in the sentence is covered into the word vector. The experimental results show that the two-way training language model has a deeper understanding of the context than the one-way language model. Among them, the text classification , Using Transformer's Encoder.

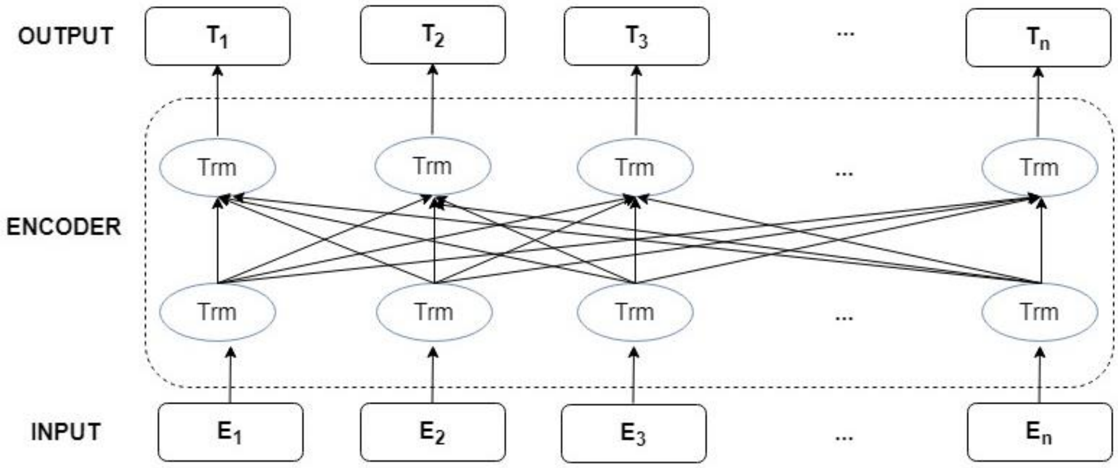

Fig. 1 BERT's Encoder

2.2 Transformer′s Encoder

The most important part of the BERT model is the bidirectional Transformer coding layer, which extracts text features using Transformer's Encoder feature extractor[4]. Encoder is composed of self-attention mechanism and feed fordward neural network[5-6]. Its structure is shown in (Fig.2). The core of Encoder is self-attention, the relationship between dozens or even hundreds of words can still be found, so that the left and right context information of each word can be fully tapped, so that you can get a bidirectional representation of the word.

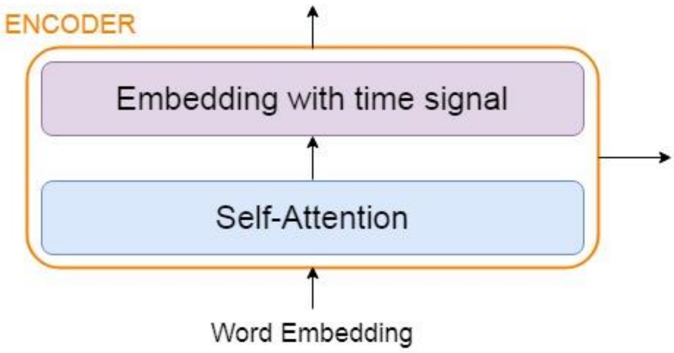

Fig. 2 Transformer's Encoder

3. Sentiment Analysis Model based on BERT

In this section, we will discuss how to perform Sentiment analysis of e-commerce comments by using the SA-BERT (Sentiment Analysis on BERT) model. Firstly, we preprocess the text data, and then send the preprocessed text to the SA-BERT model for training. See Fig 3.

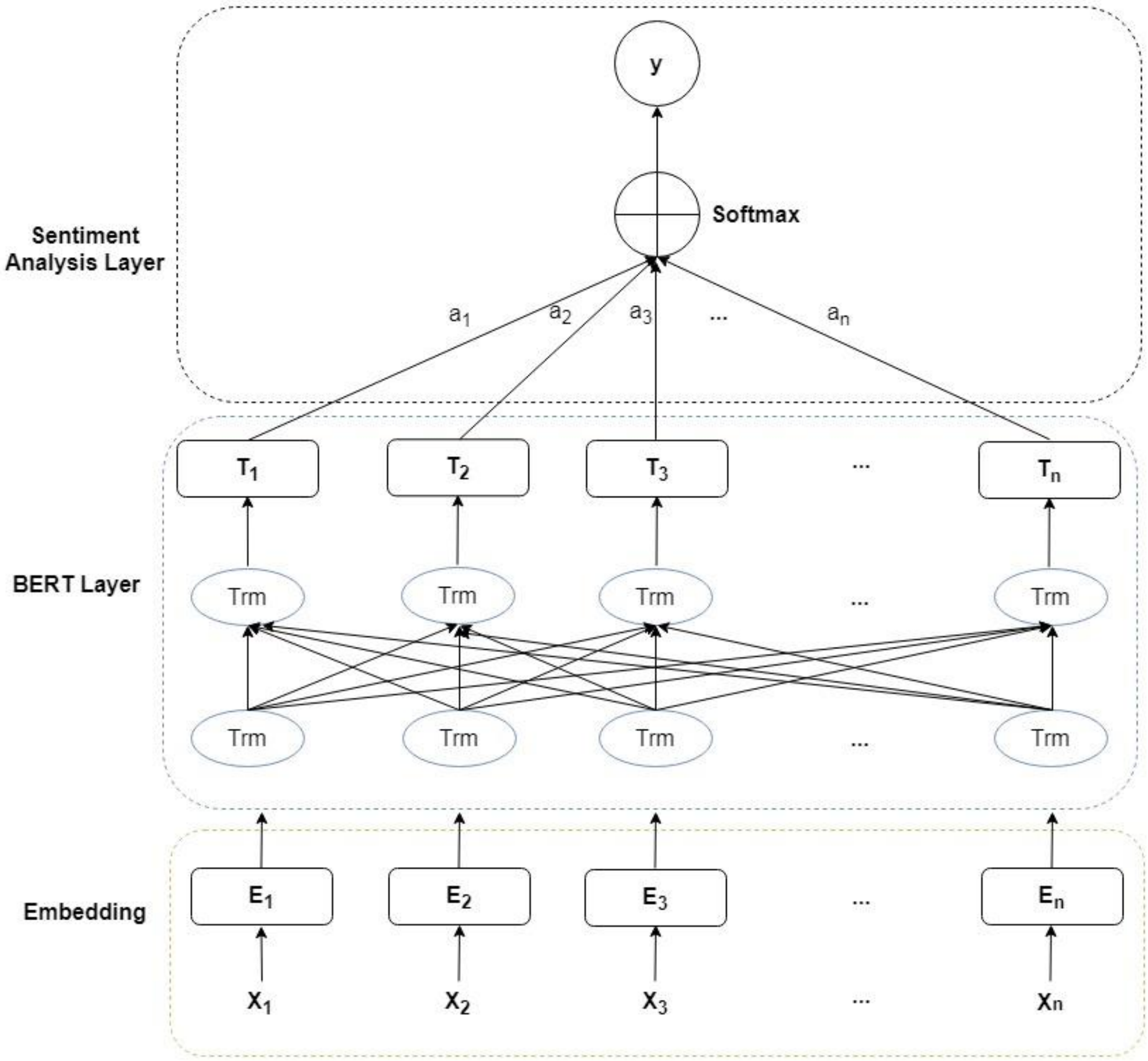

Fig. 3 Sentiment Analysis Model Based on SA-BERT

3.1 Data preprocessing

Due to the characteristics of the neural network, it is stipulated that the input text vector group must have the same dimension, so the text needs to be processed. After analyzing the text, it is found that the text with a length of more than 50 Chinese characters (one word in English is one Chinese character) accounts for 6.7% of the total text. In order to retain the text information while improving the training efficiency and making the data meet the requirements of the model, set the length of a single text to 50 Chinese characters. The part of the text with more than 50 Chinese characters is discarded, and the text with insufficient text length is filled with #.

3.2 SA-BERT Model

SA-BERT model is mainly composed of 3 parts (Embedding, BERT layer, Sentiment Analysis layer). The model structure is shown in Fig 3. The overall process of the model is: first, we use the Bert pre training model to obtain the word vector containing context semantic information, and then use the attention mechanism to extract important context related features, information, and Sentiment polarity of classified text.

3.2.1 Word Embedding

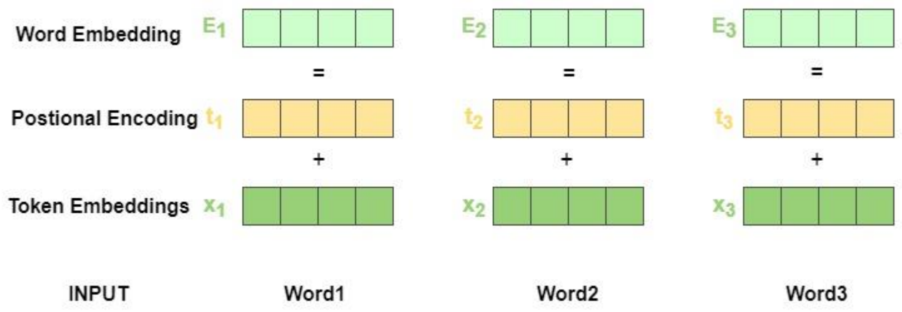

Fig. 4 The position embedding of BERT

The position embedding (shown in Fig.4) can represent the sequence of sentence sequence information, which has an important influence on the model learning sentence meaning. Transformer uses a sine wave to calculate the position information of the token, which is similar to the periodic change of the analog signal. Such a cyclic function can increase the generalization ability of the model to a certain extent. The calculation formula is as follows:

$$\mathrm{PE}_{2i}(p) = \sin(pos/10000^{2i/dpos}) \tag{1}$$

$$\mathrm{PE}_{2i} + l(p) = \cos(pos/10000^{2i/dpos}) \tag{2}$$

Map the position with id p to a d-dimensional position vector. The value of the i-th element of this vector is $\mathrm{PE}_i(p)$. However, BERT directly trains a position embedding to retain position information. Each position initializes a vector randomly and joins the model for training. An embedding containing position information can be obtained. Finally, the position embedding and token embedding are directly stitched into word embedding.

In addition, BERT uses a masked language model to achieve a true bidirectional encoding, randomly masked the expected 15% of the token, and then send the final hidden layer vector output from the masked token position to the classifier to predict the masked token Similar to the cloze, although you can see all the location information, but the words that need to be predicted have been replaced by special symbols and can be directly encoded. But after confirming the masked words, they are not directly removed, but 80% will be directly replaced, 10% will be replaced with any word, and 10% will retain the original token. This is to enhance the robustness of the model, avoid words that the model does not recognize, and enhance its generalization ability.

3.2.2 BERT Layer

The steps of generating word vectors based on the traditional deep learning text sentiment analysis method are: word segmentation, removing stop words, and generating word vectors[7]. The most commonly used word vector is Word2Vec, but it cannot capture the bidirectional semantic features of the text well. Because BERT makes full use of the information on the left and right sides of the word to obtain a better distributed representation of the word, this text will send the text into the BERT to generate the word vector instead of the word vector generated by Word2Vec.

The input of the BERT model is the sum of two vectors. These are the word vector (token embedding) and the position vector (Position Embedding). The word vector indicates the encoding of the current word, and the position vector indicates the encoding of the current word position. Each sentence uses <CLS> and <SEP> as the beginning and ending markers.

In the experiment of this paper, the pre-trained Chinese model "BERT-Base, Chinese" published by Google is used. The model uses a 12-layer Transformer with a hidden size of 768, a Multi-head Attention parameter of 12, and a total model size of 110MB. After loading the model, you can directly output the trained word vector or sentence vector. This article uses this model to obtain the sentence vector and use it as the input of the subsequent network model.

3.2.3 Sentiment Analysis Layer

Finally, the output of the BERT layer goes through the Sentiment Analysis layer, and when the vector representation of the sentence is obtained, different weights are given to different words in the comment text, and then the vector representation of the sentence is weighted by these different weighted word vectors.

$$a_i = \frac{\exp(e_i)}{\sum_{j=1}^{n} \exp(e_j)} \tag{3}$$

$$e_i = \tanh(Wh_T + b) \tag{4}$$

Which a_i represents the attention probability weight of node i. W represents the weight matrix, h_T is the output of BERT.

Then the fully connected layer integrates the extracted features to perform a fully connected operation. Different weights are given to different word vectors, then softmax is normalized, and finally the corresponding category is output through the defined discriminant function.

4. Experiments

This section compares the SA-BERT model with several other models on the JingDong mobile phone review data set. It not only compares the accuracy and stability of the model, but also compares the time-consuming degree of model training.

4.1 Experimental Data

The data set used in this article is JingDong mobile phone review data, a total of 22087 data. Among them, there are 11425 positive evaluation data and 10662 negative evaluation data,and half of them will be randomly trained data C_{train}, half as the test data C_{test}.

4.2 Experimental setup

In SA-BERT model, the text length is 100, and the word vector dimension is 768; the LSTM hidden layer has 64 neurons, Adam's learning rate is 0.001, and dropout is 0.5. The experiment compared the accuracy and recall rate of the AT-LSTM-BERT model with the three methods of LSTM,AT-LSTM, and BERT.

LSTM: Using word2vec as the word vector input, the word vector dimension is 768; among them, the LSTM hidden layer has 64 neurons, Adam's learning rate is 0.001, and dropout is 0.5.

AT-LSTM: Same as LSTM model.

BERT: The pre-trained Chinese model "BERT-Base, Chinese" published by Google. The model uses a 12-layer Transformer with a hidden size of 768, a Multi-head Attention parameter of 12, and a total model size of 110MB.

4.3 Experimental Results

Compare the AT-LSTM-BERT model with several other comparison models(LSTM,AT-LSTM, BERT) in the table 1. It can be seen that compared with these commonly used models, the SA-BERT model has a significant improvement in accuracy and F value, which shows that SA-BERT is on the Jingdong mobile phone review data set , The effect of sentiment classification is better. At the same time, compared with using LSTM or AT-LSTM alone for text classification, it has also been significantly improved.

Table 1 Experiment results

Model	Accuracy	Precision	Recall	F1
LSTM	0.8141	0.8412	0.8573	0.8492
AT-LSTM	0.8771	0.8845	0.8652	0.8747
BERT	0.9062	0.9153	0.8876	0.9012
SA-BERT	0.9135	0.9481	0.9153	0.9314

Comparing the time used by the SA-BERT model and other models in Table 2, it can be seen that except for the BERT model, although the accuracy of other models is relatively lower than the F value, the time used is obvious compared to the mixed model SA-BERT A lot less, which shows that the BERT pre-trained word vector model requires a lot of time to represent the word vector in the sentence. SA-BERT works well, but it also takes a lot of time.

Table 2 Running time

Model	Time
LSTM	0.74
AT-LSTM	0.87
BERT	5.70
SA-BERT	5.80

5. Concluding

In recent years, the development of e-commerce has become more and more popular, and sentiment analysis of e-commerce reviews has become a research hotspot. For better research, this paper proposes a SA-BERT network combining BERT word vector and attention mechanism. This model is used for e-commerce review research and analysis. The model first uses BERT to encode word vectors to represent the semantic information of the text, and then uses the attention mechanism to extract text features in a deeper way to deeply understand the semantics of the text information and complete sentiment analysis of e-commerce review. Use mobile phone review data The simulation experiment is carried out on the set, and the experimental results show that the AT-BERT model greatly improves the performance of sentiment classification.

Acknowledgements

This work is partially supported by National Natural Science Foundation of China (81973981) and General soft science project of Shandong Province in 2019.

References

[1] Tang D, Qin B, Liu T. Document Modeling with Gated Recurrent Neural Network for Sentiment Classification[C]. Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing. 2015.

[2] Hochreiter S, Schmidhuber J. Long Short-Term Memory[J]. Neural computation, 1997, 9(8): 1735-1780.

[3] Devlin J, Chang M, Lee K, et al. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding[J]. arXiv: Computation and Language, 2018.

[4] Vaswani A, Shazeer N, Parmar N, et al. Attention is All you Need[C]. neural information processing systems, 2017: 5998-6008.

[5] Devlin J, Chang M, Lee K, et al. BERT: pre-training of deep bidirectional transformers for language understanding[C]. North American Chapter of the Association for Computational Linguistics, 2019: 4171-4186.

[6] Ma Y, Peng H, Cambria E. Targeted aspect-based sentiment analysis via embedding commonsense knowledge into an attentive LSTM[C]. Thirty-Second AAAI Conference on Artificial Intelligence. 2018: 5876-5883.

[7] Huang B, Carley K. Parameterized convolutional neural networks for aspect level sentiment classification[C]. Empirical Methods in Natural Language Processing, 2018: 1091-1096.