Advances in Artificial Intelligence, Big Data and Algorithms
G. Grigoras and P. Lorenz (Eds.)

doi:10.3233/FAIA230903

The Research on Aspect-Level Sentiment Analysis for Online Reviews Based on BERT-BiLSTM Model

Haiqin TANG[1] and Ruirui ZHANG
Business and Tourism school, Sichuan Agricultural University, China

Abstract. With the development of professional online review systems, current sentiment analysis research is more focused on classifying the sentiment polarity of different aspects of the same subject, which also helps consumers to make decisions. To address the problems of multiple meanings of words, inadequate contextual semantic understanding and incomplete feature extraction in online reviews, this paper proposes an aspect-level sentiment analysis model based on BERT-BiSTM (Bidirectional Encoder Representation from Transformers and Bi-directional Long Short-Term Memory). The BERT word vectorisation can make use of the information in both directions in the text and can better learn the semantic information of the text. The results of the BERT model are fed into the BiLSTM model to fully perform feature mining for aspect extraction and sentiment polarity classification. The experimental results show that the model in this paper works well on the tasks of aspect extraction and sentiment polarity classification on the four datasets Semeval2014_restaurant (Res14), Semeval2014_laptop (Lap14), Semeval2015_restaurant (Res15) and Semeval2016_restaurant (Res16).

Keywords. BERT pre-training, BiLSTM, Online reviews, Aspect-level sentiment analysis

1. Introduction

Sentiment analysis in natural language processing refers to the use of computer-aided tools to analyse people's subjective feelings such as opinions, emotions, evaluations, perceptions and attitudes towards physical objects such as products, services, individuals, events, topics and their attributes, based on text. Aspect-level sentiment analysis is one of the most important current concerns in the field of sentiment analysis methods and consists of two sub-tasks: the aspect extraction task and the aspect-level sentiment classification task. The aspectual word can be a word or a phrase.

Aspect and sentiment analysis techniques are used in many fields and are being used and researched with great enthusiasm in the field of e-commerce. With the widespread use of artificial intelligence and big data technologies, reviews on online shopping sites contain huge potential value that can help consumers' choices and companies' decisions in the product development and sales process.

In the age of information explosion, online comments on websites are updated daily in real time and cover a large amount of data, posing a huge challenge to sentiment

[1] Corresponding Author, Haiqin TANG, Sichuan Agricultural University, China;
Email: 1696562280@qq.com

analysis techniques. Through the analysis of comment texts and the comparison of classification models. Hu and Liu [1] were the first to propose the use of unsupervised methods, and they extracted commodity features by association rule mining methods. Popescu et al. [2] introduced Point Mutual Information (PMI) to identify evaluation and non-evaluation objects, which improved the recognition of low-frequency words. An online review may contain different aspects of sentiment, so when we processing the data and extracting aspect words, it is important to not only ensure efficient extraction, but also to focus on obtaining semantic information from the text. In addition, there are often short sentences or sentiment polarities that are not clearly expressed in user comments, which have limited sentiment features and are not well identified by independent sentence input. The BERT model can make use of the information in both directions of the text to better learn the semantic information of the text, and the Bert model has significant effect on the processing of large amount of data. Secondly, this paper incorporates the BiLSTM model for sentiment classification, which can better capture the semantic dependencies in both directions and obtain the dependencies of specific aspects in the sentence and in the whole comment, and improve the classification effect of the model.

2. Related Works

Aspect-level sentiment analysis is carried out by inferring the sentiment polarity of aspects of a text given a passage and a number of aspects of that passage. Sentiment polarity is generally classified as positive, negative or neutral.

2.1. Aspect Extraction

Aspect word extraction refers to the extraction of nouns or noun phrases with different attributes or features of the product being reviewed in a sentence of a product review, so aspect word extraction is a sequential annotation problem.

Jakob et al.[3] selected a variety of features such as words, lexicality, dependent syntax, etc., and used CRF models to extract evaluation features from movie review data, achieving better results. liu et al.[4] attempted to achieve the extraction of sentiment features from evaluation texts by training a deep learning neural network model (RNN); Poria et al.[5] attempted to automatically extract feature information from text by training a deep learning convolutional neural network; Xin[6] by integrating convolutional neural networks and convolutional recurrent neural networks, using the two networks to train a novel model for text feature extraction. Li et al.[7] extracted evaluation objects and evaluated sentiment words via LSTM models, while demonstrating that this method performed better than traditional rule-based bidirectional propagation algorithms. Zheng et al.[8] used CNN to extract contextual short-distance features and BiLSTM to extract long-distance hidden features, and then spliced the two features by the concept of location weight information; Li et al.[9] used a BiLSTM model to extract contextual short-distance features and a BiLSTM model to extract long-distance hidden features.

2.2. Sentiment Classification

After extracting the aspect words of the online reviews, it is necessary to classify their opinions based on the aspect words in terms of polarity, and this part of the task is similar to sentence-level sentiment analysis, which can also use sentiment lexicon based, machine learning based and deep learning based classification methods. Hu and Liu [1] proposed a sentiment lexicon based approach for sentiment determination. Although sentiment lexicons have achieved certain results in sentiment determination, this unsupervised lexicon based approach suffers from poor domain applicability and low accuracy, so supervised sentiment classification methods have been widely studied and applied. As research progresses, a large number of scholars try to apply deep learning methods to the field of text sentiment analysis and achieve good results. Han et al [10] firstly used CNN model to obtain local features of text, then used LSTM to obtain sentence-level semantic representation in text, then fused user features and product attributes with semantic representation, and finally input to another LSTM classification model to finally achieve sentiment classification of text. Xue [11], Zhang[12] et al. effectively improved the performance of the model for sentiment analysis by considering the interdependencies between words through bi-directional gated recurrent units (BiGRU). lv et al.[13] proposed a simple and effective joint sentence length-based aspect sentiment analysis framework by combining the needs of two subtasks, aspect-oriented term extraction and aspect-oriented sentiment classification, in aspect-level sentiment analysis, which outperformed other models on three benchmark datasets. yang et al.[14] In response to the inability of existing models to correctly establish word-to-word prior dependencies, the word-implicit vector of LSTM output was first combined with the dependency embedding to form a combined vector.

2.3. BERT

The BERT model is a pre-trained model proposed by Google in 2018, and the use of the BERT model has facilitated research in sentiment analysis. Kang et al.[15] took advantage of BERT pre-training to extract commodity entities, attributes and sentiment words using a BILSTM-CRF attention mechanism model fusing syntactic features with BERT word embedding, and used BILSTM to perform sentiment analysis on the extracted results, which outperformed the BILSTM model and BILSTM-CRF model alone. Yu et al.[16] used a BERT pre-training model to generate feature vectors fusing contextual semantics and aspectual words, combined with a multi-headed self-attentive mechanism, and used gated convolutional networks for aspect-level sentiment classes, and demonstrated the improvement in accuracy of the model through different experiments. Li et al.[17] used the phrase initialization vector output from BERT as a mnemonic to interact with the vector of aspectual phrases with full attention, and the vector formed by splicing was used as the final sentiment classification feature vector with better results compared with other models. Zeng et al.[18] designed a label that contains both aspectual word location information. A pre-trained language model, BERT, was used to automatically learn feature representations with contextual information. Yang et al.[19] used BERT to obtain node information containing contextual and aspectual word interaction attention; meanwhile, a gated graph neural network was used to update the nodes, after which the global information was further fused in the self-attention layer.

For aspect-level sentiment analysis, both aspect extraction and aspect-based sentiment classification, deep learning models are preferred, the BERT model is a popular pre-training model that is widely used and effective for both feature extraction and sentiment classification tasks. Improved models such as Elmo, GPT, Roberta, etc. can also be used for different tasks and scenarios. After using the BERT model for aspect extraction, the BiLSTM neural network is used for sentiment classification to improve the overall effectiveness of the model.

3. BERT-BiLSTM

The BERT-BiLSTM aspect-level sentiment classification model proposed in this paper first uses the BERT model for word vectorisation, uses the results as input to the BiLSTM model, and finally connects a classification layer for sentiment classification. The model consists of an input layer, a BERT embedding layer, BiLSTM feature extraction and a classification output layer. The model diagram in this paper is shown in figure 1:

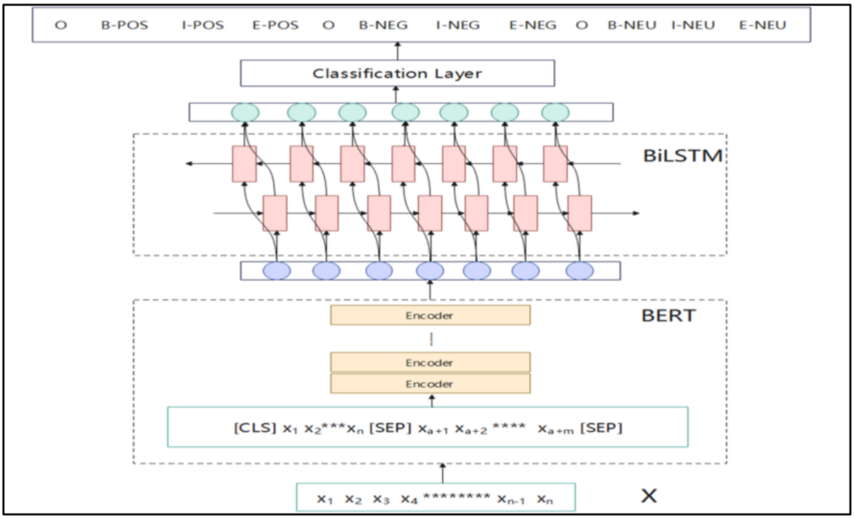

Figure 1. BERT-BiLSTM model.

3.1. Input Layer

Given a text sequence meaning $X = \{x_1, x_2,...,x_n\}$, where x_i is the i-th word in the text and n is the length of the text sequence. The aspect words are the words that appear in the text sequence, the aspect words can be represented as $X_a = \{x_{a+1}, x_{a+2}, ..., x_{a+m}\}$, where $1 \leq m \leq n-a$, the length of the aspect word is m, a+1 is the position of the beginning of the aspect word, $a \geq 0$. The input layer data in this paper is input by using the "[SEP]" notation to splice the context and the aspect word, expressed as "[CLS] + context + [SEP] + aspect word + [SEP]".

3.2. Embedding Layer

In the field of natural language processing, common word vectorisation methods include Word2Vec, Glove, etc. However, word vector models such as Word2Vec and Glove cannot cope with multiple meanings of words. The BERT pre-training model can be used to learn the semantic information of the text by using the information in both directions, so we choose the BERT pre-training model in the embedding layer.

The input vector of BERT consists of the sum of Token Embedding, Segment Embedding and Position Embedding, and this representation can This representation can solve the problem of multiple meanings of words. Based on the target text, we choose the bert-base-uncased model of BERT as the pre-training model. After the processing of the BERT model, the output of the last layer of transformer encode, H^l, is used as the input of the subsequent model.

$$H^l=\mathrm{Transformer}_l(H^{l-1}) \tag{1}$$

3.3. BiLSTM Layer

The BiLSTM is composed of two inverted LSTMs above and below, and the output of the whole network is jointly determined by the two LSTMs, making up for the inability of the LSTM to encode information from back to front.

The forward long and short-term memory network computes a sequence of implicit states of the same length as the sentence for the forward input words:$\vec{H}$ = $\{H^1{}_L, H^2{}_L, ..., H^t{}_L\}$, while the reverse LSTM computes the same implicit state sequence as the sentence length for the reverse input: $\overleftarrow{H}$ = $\{H^t{}_R, H^{t-1}{}_R, ..., H^1{}_R\}$. Finally, the vectors of the forward and reverse hidden layer state outputs are concatenated to obtain the final output of the BiLSTM network layer H = $\{\vec{H}\ \overleftarrow{H}\}$. The equations of H of the time of t are shown as follow:

$$\vec{H}_t=\mathrm{LSTM}(H_t, \vec{H}_{t-1}) \tag{2}$$

$$\overleftarrow{H}_t=\mathrm{LSTM}(H_t, \overleftarrow{H}_{t-1}) \tag{3}$$

$$H_t=w\vec{H}_t + v\overleftarrow{H}_t + b_t \tag{4}$$

3.4. Classification Output Layer

The output of the BiLSTM feature extraction layer is fed into the output layer to obtain the results of the sentiment classification corresponding to the aspect words. The equation of h^* is below.

$$h^* = \mathrm{Softmax}(H) \tag{5}$$

4. Experiments and Analysis of Results

4.1. Experimental Data

There are four validation model datasets in this paper, namely the laptop and restaurant datasets from Res14, Lap14, Res15 and Res16, and the number of sentences in the training, validation and test sets for each dataset were divided in the ratio of 8:1:1. The sample label distribution of the datasets is shown in the table 1.

Table 1. The sample label distribution of the datasets

dataset	Positive	Negative	Neutral	Total
Lap14	987	866	460	2313
Res14	2164	805	633	3602
Res15	1198	403	53	1654
Res16	1578	205	217	2507

4.2. Experimental Parameters and Evaluation Indicators

The model in this paper was programmed in Python and implemented in the Pytorch deep learning framework. And we choose bert-base-uncased as the BERT pre-training model. The parameters of the model were set as shown in the table 2:

Table 2. The parameters of the model

Parameter name	Parameter values
Max_seq_len	128
Batch_size	8
Learning rate	5e-5
Optimizer	Adam
Loss	focal
Dropout rate	0.1

Macro_F1 was chosen as the evaluation metric for the model. Macro_F1 is obtained by calculating the F1 value for each category separately on each confusion matrix now and then averaging the values in the region. The formula of Macro_F1 is as follows:

$$\mathrm{Macro\_F1} = \frac{1}{n}\sum_{i=1}^{n} \mathrm{F1}_i \tag{6}$$

4.3. Analysis of Experimental Results

4.3.1. Model Comparison Experiments

In order to verify the accuracy and feasibility of the models proposed in this paper, the following models were selected for comparison tests. Provided that the parameters are all kept consistent, the experimental results data of the models in this paper and the comparison models in the four data sets are shown in the table 3:

Table 3. The result of different models

Model	Macro_F1 (%)							
	Res14		Lap14		Res15		Res16	
	Aspect	Polarity	Aspect	Polarity	Aspect	Polarity	Aspect	Polarity
BERT-Linear	92.29	73.80	85.81	72.48	89.10	59.19	85.30	67.15
BERT-Crf	90.80	70,71	86.04	77.98	88.67	53.03	85.73	68.12
BERT-Self_attention	90.66	67.84	85.11	75.00	86.27	56.15	85.36	67.13
BERT-BiLSTM	90.98	75.38	86.33	77.99	88.84	77.27	85.57	70.83

The results in the above table show that the BERT-BiLSTM model performs better on both the aspect extraction and sentiment classification tasks in most of the experimental datasets. Compared with other models, the BERT-BiLSTM model is able to perform word vectorisation faster and solve the problem of multiple meanings of words. The BiLSTM model fully incorporates sentence contextual information and effectively extracts sentence features, thus improving the accuracy of aspectual word recognition and sentiment polarity classification to some extent.

4.3.2. Pre-training Model Selection

In this paper, the Bert model is used for pre-processing in the process of word vectorization. In order to verify the effectiveness of the Bert model, the article takes the BERT-BiLSTM model as the baseline model with the same downstream model and parameter settings.We set Bert, glove and Elmo as pre-training models, respectively, the results of each model Macro-F1 are shown in the following figure 2:

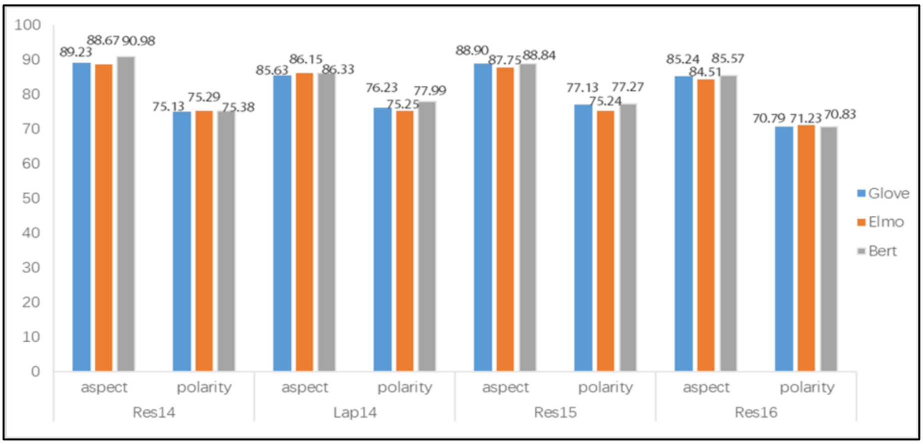

Figure 2. The result of different pre-training models.

From the above figure, it can be seen that the based-aspect sentiment analysis model combined with BERT works better than the two models using Glove and Elmo in the comment aspectual sentiment analysis task. Probably because the dynamically encoded word vector model can differentially characterise synonyms based on contextual information, solving the problem of multiple meanings of the word existing in synonyms in the comment text.So the BERT model is chosen in this paper as the word vector model.

4.3.3. BiLSTM Layer Selection

In order to investigate the influence of the number of layers of LSTM on the accuracy of the model in the BiLSTM model and to determine the optimal number of model layers. We set the number of LSTM layers as 1, 2, 3 and 4 respectively on the basis of the BERT-BiLSTM model, and the final results of the model with different numbers of layers are as follows:

As shown in the table 4, when the number of LSTM layers is 2 or 3, the F1 of the model is relatively high, and when the number of layers is greater than 3, the effect of the model starts to decrease and the experimental time consumption increases accordingly. The bottom LSTM generally characterises literal information, while the deep LSTM generally characterises semantic information. Therefore, it is necessary to increase the number of LSTM layers, so we choose 2 as the number of the LSTM in this paper.

Table 4. The result of different Number of LSTM layers

Number of LSTM layers	Macro_F1 (%)							
	Res14		Lap14		Res15		Res16	
	Aspect	Polarity	Aspect	Polarity	Aspect	Polarity	Aspect	Polarity
1	88.23	69.53	86.27	80.12	87.28	77.21	84.29	70.58
2	90.98	75.38	86.33	77.99	88.84	77.27	85.57	70.83
3	90.24	76.36	85.21	77.26	88.79	76.38	85.39	69.79
4	89.25	75.21	85.14	76.31	86.39	75.21	85.40	70.02

5. Summary

In aspect-level sentiment analysis, in order to solve the problems of multiple meanings of words in online comment sentences and inadequate semantic understanding and feature extraction, this paper proposes a sentiment analysis model based on BERT-BiLSTM. In which BERT and the training model fully understand the semantics through dynamic coding, and BiLSTM utilizes bi-directional LSTM layers to fully extract contextual features. The experimental results show that the model in this paper works well in aspect extraction and sentiment polarity classification for most of the datasets, but there is still room for further improvement of the model, such as modules that can be built to incorporate lexical features. Most of the current research objects are online reviews in English, so further research can be conducted on aspect-level sentiment analysis of online reviews in Chinese text or mixed text in multiple languages.

References

[1] Hu.M, Liu.B, Mining opinion features in customer reviews [A]. *Proc of 19th National Conference on Artificial Intelligence*[C]. Palo Alto:AAAI Press, 2004, pp: 755-760.
[2] Popescu A M, Etzioni O. Extracting product features and opinions from reviews [M]. *Berlin:Springer*, 2007:9-28.
[3] Jakob N, Gurevych I. Extracting opinion targets in a single- and cross-domain setting with conditional random fields [A]. *Proc of the 2010 Conf on Empirical Methods in Natural Language Processing. stroudsburg*[C]. Pennsylvania: ACL, 2010:1035-1045.
[4] Liu P, Joty S, Meng H. Fine-grained Opinion Mining with Recurrent Neural Networks and Word Embeddings [A]. *Proc of the 2015 Conference on Empirical Methods in Natural Language Processing*[C]. Stroudsburg: Association for Computational Linguistics, 2015:1433-1443.
[5] Poria S, Cambria E, Gelbukh A. Aspect Extraction for Opinion Miningwith a Deep Convolutional Neural Network[J]. *Knowledge Based Systems*, 2016,108: 42-49.
[6] Wang X.Y. A deep learning method for text feature extraction[J]. *Microcomputer Applications*,2021,37(01):137-139.
[7] Li S Q, Zhao Y Y, Qin B ,Liu T. Evaluation object and evaluation word extraction based on LSTM network [J]. I*ntelligent Computers and Applications*,2017,7(05):95-97.
[8] Zheng C, Cao Y, Xue X. A multilayer attention network for aspect-level sentiment classification[J]. *Computer Engineering and Applications*,2020,56(19):176-181.
[9] Li J Y, Kang Y, Yang Q Y, Wang P Y, Cui G R. Fine-grained sentiment analysis of IT products for aspect-oriented memory networks[J]. *Computer Engineering and Applications*,2020,56(03):159-164.
[10] Han H, Liu G L, Sun T Y. Multi-attentive hierarchical neural network for text sentiment analysis[J]. *Computer Engineering and Applications*, 2020, Vol. 10:105-110.

[11] Lai X M, Tang H, Chen H Y. Multimodal sentiment analysis of feature fusion-bi-directional gated cyclic units based on attention mechanism[J/OL]. *Computer Application*s:1-9.
[12] Zhang J Q. CNN and BiGRU-attention based method for identifying sensitive entities on the Internet[J]. Network Security Technology and Applications, 2020, Vol. 4:61-65.
[13] Lv Y, Wei F, Zheng Y. A span-based model for aspect terms extraction and aspect sentiment classification [J]. *Neural Computing and Applications*, 2020(2).
[14] Yang Z, Li C, Zhao Z. Sentiment classification based on dependency relationship embedding and attention mechanism[J]. *Journal of Intelligent and Fuzzy Systems*, 2021, 41(1):1-11.
[15] Kang Y, Xue H Z, Hua B. Fine-grained commodity evaluation analysis for deep learning networks[J]. *Computer Engineering and Applications*,2021,57(11):140-147.
[16] Yu B G, Zhang S W. Research on aspect-level sentiment analysis based on BAGCNN[J]. *Data Analysis and Knowledge Discovery*,2021,5(12):37-47.
[17] Li P,Wu Y D, Chu Q K, Fu C S, Zhang G Y. Aspect-level sentiment analysis of long texts based on BERT and memory networks[J]. *Sensors and Microsystems*,2022,41(02):118-122.
[18] Zeng F X, Li X, Yao C L, Fan F L. End-to-end aspect-level sentiment analysis based on BERT[J]. *Journal of Dalian University of Technology*,2022,41(03):228-234.
[19] Yang C X, Yao S C, Song J J. An aspect-level sentiment analysis model based on word co-occurrence[J]. *Computer Engineering and Science*,2022,44(11):2071-2079.