Sentiment Analysis of Film Reviews Based on Deep Learning Model Collaborated with Content Credibility Filtering

Xindong You[1], Xueqiang Lv[1], Shangqian Zhang[1], Dawei Sun[2(✉)], and Shang Gao[3]

[1] Beijing Key Laboratory of Internet Culture and Digital Dissemination Research, Beijing Information Science and Technology University, Beijing, China
youxindong@bistu.edu.cn
[2] School of Information Engineering, China University of Geosciences, Beijing, China
sundaweicn@cugb.edu.cncs
[3] School of Information Technology, Deakin University, Waurn Ponds Victoria, Geelong, Australia
shang.gao@deakin.edu.au

Abstract. Sentiment analysis of film reviews is the basis of obtaining the opinions of movie viewers. It has an important influence on movie public opinion control and stimulating potential viewers. Due to the natural openness and randomness of social media, there may exist a considerable amount of useless or false information in film review comments, making it challenging to analyze the credibility of the comments. This paper proposes a fine-grained sentiment analysis method based on the key-viewpoint sentences of Chinese film reviews, where a deep learning model is used to classify the fine-grained emotions in film reviews. Based on the analysis results, a method for calculating the credibility of review comments is proposed. Under the credibility criteria, corpus screened through credibility filtering algorithm, the overall sentiment classification can obtain 9% improvement on accuracy than the original corpus, which verifies the validity of the credibility algorithm. The higher quality corpus achieved by the credibility algorithm is benefit for improving the accuracy of the sentiment classification.

Keywords: Sentiment analysis · Film reviews mining · Natural language processing · Deep learning · Credibility algorithm

1 Introduction

With the rise of social media platforms and their widespread use, natural language processing for social network data has become a hot research topic. In terms of movie review, more and more people are publishing their review comments and opinions on film forums or relevant websites. These comments and opinions contain emotions, perceptions, attitudes, feelings, and behavioral tendencies towards films, cast members, the film industry, and its market. Internet word-of-mouth has a significant impact on

H. Gao et al. (Eds.): CollaborateCom 2020, LNICST 349, pp. 305–319, 2021.
https://doi.org/10.1007/978-3-030-67537-0_19

consumers' choice of watching movies. The higher the word-of-mouth evaluation, the higher is their box office revenue. Therefore, sentiment analysis of film evaluation helps understand the emotional tendency of the audience and timely acquire their opinions and attitudes [1]. It is very important for movie public opinion control, product marketing, and stimulation of potential viewers.

Text sentiment analysis, also known as opinion mining, is the process of analyzing, processing, summarizing and reasoning subjective texts with sentimental tendencies [2]. For example: "The rhythm is just to the point". This sentence expresses a positive view. In the sentence, "rhythm" is used as the characteristic word, and the corresponding emotional words are "just to the point", then "rhythm" and "just to the point" are pairs of characteristic viewpoint words. The short sentence of the viewpoint words is called a key viewpoint sentence. It can be seen from the sentiment words that the emotion of the sentence belongs to positive emotion. This paper uses a Bi-GRU model to perform a fine-grained sentiment analysis on key-viewpoint sentences. Based on the analysis result, a credibility algorithm for film review filtering is proposed. A Bi-LSTM is used to analyze the sentiment of the filtered reviews with higher credibility.

Due to its natural openness, virtuality, and randomness, especially the communication and word-of-mouth characteristics, social media have great commercial benefits for marketing and advertising. An 'internet water army' may be hired to publish a large number of false contents, including a considerable amount of useless or false information to influence the audience. It is important to analyze the content, identify these false comments, and evaluate the comment credibility. To accomplish the above objectives, we first analyze the characteristics of emotional tendency, sentence length and position of key-viewpoint sentences. Based on the analysis result, we then propose a credibility algorithm to filter film reviews. By classifying the overall sentiment on the credibility-filtered comments, 9% accuracy improvement is obtained than that of the unfiltered original corpus, which proves the validity of our credibility algorithm.

As a whole, the main contributions of this paper are listed as bellow.

(1) Corpus for extracting the fine-grained sentiment tendency is constructed, in which key viewpoint sentences and the sentiment tendency of the key viewpoint sentences are labelled.
(2) Comparison experiments are conducted on the corpus to analyze the fine-grained sentiment tendency with different deep learning model.
(3) Comments filtering algorithm is designed for obtaining higher quality corpus for overall sentiment analysis, which is based on the fine-grained sentiment analysis, length of the key viewpoint sentence and the position of the key viewpoint sentence.
(4) Extensive experiments are conducted on the corpus filtered by the different credibility with different mainstream deep learning model, which show that 9% improvement is achieved than on the un-filtered original corpus with Bi-LSTM employed.

2 Related Work

Sentiment Analysis can be divided into coarse-grained and fine-grained by granularity. Coarse-grained sentiment analysis only judges the overall emotional tendency at the text

level. However, the overall sentimental tendency of a review text may be different or opposite to the emotional propensity of the individual evaluation opinions in the evaluation text [3]. Therefore, a finer granular analysis of sentiment is needed to identify the emotional tendencies of different opinions in the evaluation text. The aspect-level sentiment classification is a fine-grained sentiment classification task [4, 5]. The goal of this task is to infer the emotional polarity of corresponding evaluation words for the evaluation objects appearing in a given text [6–8]. Fine-grained sentiment analysis focuses on the extraction and analysis of attribute words and emotional words. Early scholars often used traditional machine learning and rule-based methods to extract emotion evaluation units in texts to perform fine-grained sentiment analysis. Liu Li et al. pruned the grammar tree by establishing a syntactic path library [9]. For texts containing multiple evaluation objects and evaluation words, they removed irrelevant evaluation objects regularly. The result of fine-grained sentiment analysis was obtained by extracting the correct evaluation unit. Tang et al. tried to improve the accuracy of microblogging fine-grained sentiment analysis by constructing a product feature ontology and extracting implicit product features by using feature word-emotional word pairs [10]. Traditional machine learning or rule methods only consider shallow semantic in-formation. Although this method is relatively simple and straightforward, it relies extremely much on information such as feature engineering and emotional dictionary. It requires knowledge support in related fields and the classification effect is not satisfactory.

Since it expresses textual semantic information with low-dimensional dense vectors, deep learning method contains deeper semantic information, which is convenient for mining and analyzing the relationship between words. Therefore, it has become a mainstream method for natural language processing related research. The TD-LSTM and TC-LSTM models proposed by Tang et al. enhanced the topic information based on the LSTM model, thereby improved the accuracy of sentiment classification for specific topics [11]. Ruder et al. implemented a hierarchical LSTM to model the correlation be-tween sentences in perspective sentiment analysis [12]. The hierarchical network structures worked better on multiple domains and language samples than models without levels. Ma et al. proposed an interactive attention network model to learn the target word and context representation separately, which completed the target word sentiment analysis task [13]. While paying attention to the overall text, they also introduced the characteristics of target words into the model. Xue et al. proposed a model based on the combination of CNN and gate mechanism [14]. The proposed structure is much simpler than the attention layer applied to the existing models, and the calculation of the model is easy to parallelize in training. Guan et al. used the attention mechanism to learn the weight distribution of each word on the sentiment tendency of the sentence directly from the word vector and used Bi-LSTM to learn the semantic information of text, and then improved the classification effect through parallel fusion [15]. Zhang et al. proposed a method of sentiment analysis of film reviews based on multi-feature fusion, which improved the sentiment analysis results on Chinese film reviews [16]. Deep learning has an excellent self-encoding ability and can well establish the mapping from low-level signals to high-level semantics. In this paper, a deep learning model Bi-GRU is selected

as the basic model in the sentiment analysis experiments at the aspect-level, and a Bi-LSTM is selected as the basic model in the sentiment analysis experiments at the text level. Improvements obtained verify the effectiveness of the models.

It is important to evaluate the credibility of the comments. At present, the research on the quality of network information is more about the exploration of network opinion leaders. In essence, it is a method of judging high-quality users, mainly relying on external indicators such as citation, forwarding amount, and comment number to conduct credibility calculation [17]. This kind of method subjectively assumes that the content published by high-quality users is of high quality. While the number of high-quality users themselves is relatively few, the quality of content published by most users cannot be detected. Sheng et al. used Weibo users who had larger forwarding numbers and comments to build a domain-related vocabulary to calculate content similarity and evaluate the credibility of published contents [18]. This method essentially excavated the shallow mining of information, but did not fully consider the deep semantic information of the text, and it relied heavily on indirect information. Inspired by the afore-mentioned work, this paper considers three important characteristics of movie review comments: emotional tendency, sentence length and position of key viewpoint sentences. By considering the result of the fine-grained sentiment analysis text-level sentiment classification is conducted on these more credible comments filtered by the review credibility algorithm. The validity of the credibility algorithm is verified by substantial experiments.

3 Sentiment Analysis Based on Content Credibility Filtering

Figure 1 is a flow chart of the proposed fine-grained sentiment analysis based on the key viewpoint sentences. It consists of 4 main parts: corpus preprocessing and key viewpoint sentence labelling part, fine-grained sentiment classification model construction part, film review credibility algorithm part, and overall sentiment classification model part.

3.1 Corpus Preprocessing and Key Viewpoint Sentence Annotation

The experimental corpus adopted in this paper is the film reviews on 2017 China Film Lists on Douban website, covering comedy, thriller, horror, crime, action and other film genres. The film reviews are divided into 1 to 5 stars according to the star rating, corresponding to five rates of emotional tendency: highly recommended, recommended, just fine, poor, bad. There are 3755 movie reviews that have been manually labelled. The length of a review is the number of words included. The average sentence length is 47.465 words.

Corpus Samples are shown in Table 1, where the first column is the corpus number, followed by the emotional tendency rating score of the original corpus, the original corpus details, the manually labelled key viewpoint sentences and the emotional tendency of key viewpoint sentences, respectively.

According to the previous research work and our analysis, key viewpoint sentences can be divided into the following 3 categories.

Category 1: Sentence expression is an emotional tendency toward a single explicit evaluation object. That sentence contains a single explicit evaluation object and corresponding context-free emotion words. For example, the sentence “a” of the above

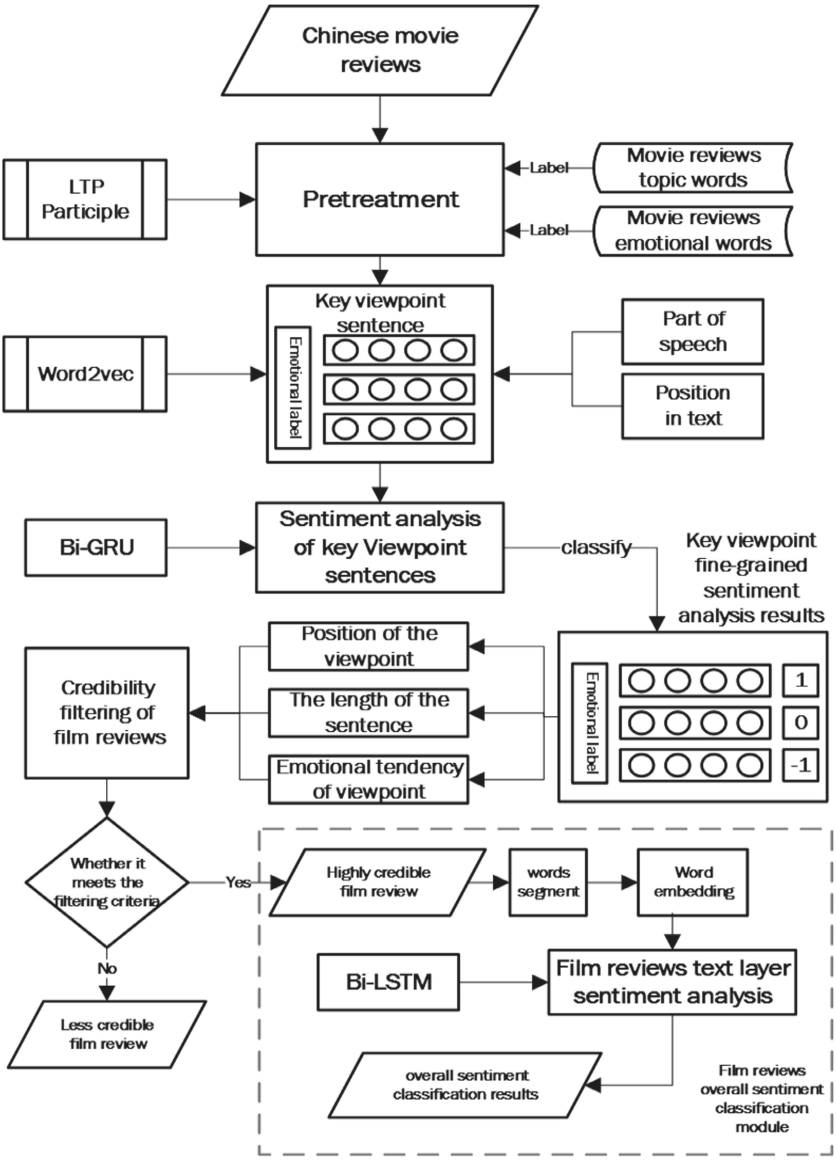

Fig. 1. Flow chart of sentiment analysis of film reviews based on content credibility filtering

example 3: "The storyline is a half-wit", this viewpoint sentence contains one single evaluation object "storyline" and the corresponding derogatory emotional tendency word "half-wit". The sentiment orientation of this sentence is "negative".

Category 2: Sentence expression is an emotional tendency toward multiple explicit evaluation objects. That sentence contains a plurality of explicit evaluation objects and corresponding context-free emotion words. For example, the sentence "a" of the above

Table 1. Corpus samples

No	Score	Original comment	Key viewpoint sentence	Emotional tendency
1	4stars	Music and filming are very good, but the story is disappointing	**a.** Music and filming are very good **b.** the story is disappointing	Positive; Negative
2	3stars	Very good, it is OK as a whole, many connecting points are sore thumb, the screenwriter has done a little bit of work, laughing points are very good, it is surprising there are cry points, I have cried twice. Unexpectly, Lu Zhengyu likes making bitter Qianlong Childhood!…	**a.** Very good **b.** it is OK as a whole **c.** many connecting points are sore thumb **d.** the screenwriter has done a little bit of work **e.** laughing points are very good **f.** it is surprising, there are cry points	Positive; Neutral; Negative; Positive; Positive; Positive
3	2stars	The storyline is a half-wit, the laughing point is awkward, the acting skill is boastful, it is a big problem for the actor' lines for the leading actress Guo Caijie, she can punctuate and stress the words correctly, two stars for it, the scene of the food is good.	**a.** The storyline is a half-wit **b.** the laughing point is awkward **c.** the acting skill is boastful **d.** it is a big problem for the actor' lines for the leading actress Guo Caijie **e.** the scene of the food is good	Negative; Negative; Negative; Negative; Positive

example 1: "Music and frames are very good", contains two evaluation objects "music" and "frame". The emotional tendency is derogatory, "good". The emotional orientation of this sentence is "positive".

Category 3: A sentence expresses an emotional tendency toward an implicit object. There are no obvious evaluation objects in the sentence, but they express obvious emotional tendencies. Sentence "b" in example 2, "very good", no explicit evaluation objects are mentioned, but there are emotional words - "very good". The sentiment orientation of this sentence is "negative".

According to the above analysis, it is found that there may be multiple key viewpoints in one comment, and each key viewpoint sentence may contain multiple opinions. After a statistical analysis of the 3,755 film reviews, the number of key viewpoint sentences included is 9019, and the number of views is 10,807.

3.2 Fine-Grained Sentiment Classification Model Based on Key Opinion Sentences

The overall architecture of the fine-grained sentiment classification model based on key viewpoint sentences is shown in Fig. 2. In this paper, a Word2vec word vector is used as the model input, a Bi-GRU model is used to learn the commentary dependency, and a Softmax layer is used as the output of the classification model to obtain the corresponding emotional tendency of the key opinion sentences. Over the corpus with key opinion sentences labelled and the emotion polarity of the key opinion sentences labelled. The key opinion sentences will be extracted and the categories of the key opinion sentences will be classified by the GRU neural network.

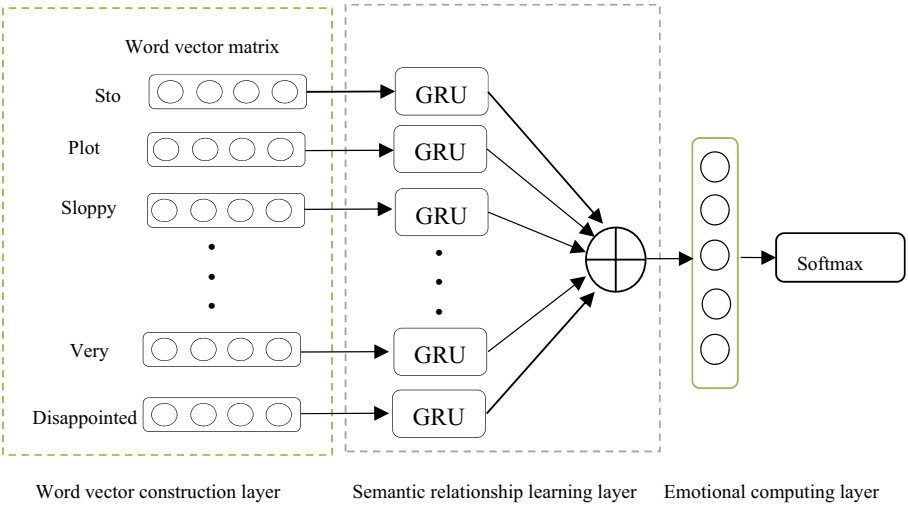

Fig. 2. Schematic diagram of the fine-grained sentiment classification model based on key viewpoint sentences.

3.3 Semantic Relationship Learning Layer

The sentiment analysis of the key opinion sentences needs to consider the semantic dependence between words. This paper uses a Bi-GRU model to obtain the semantic dependence information from film reviews. GRU (Gated Recurrent Unit) was proposed by Cho [19], which is a variant of the LSTM[20] model. It removes the forgetting gates, input gates, and output gates in the LSTM and only retains the reset gates and update gates. The advantage of LSTM model is kept but with a simplified structure.

The update gate of GRU controls how much information in the hidden state is passed to the current hidden state. The greater the value of the update gate, the more information is updated at the current time. This is very similar to the memory unit in a LSTM network, which helps GRU learn long-term dependencies. Since the training parameters are less than those of LSTM, its training speed is faster.

The Bi-GRU model is used to improve the accuracy of emotional classification results in key sentences. This structure provides complete forward and backward context information about the input sequence at the output layer.

3.4 Emotional Computing Layer

Define h_t to represent the feature of sentiment classification. Softmax classifier is used to classify key viewpoint sentences. The predicted sentiment category of a comment is y^, and y is the actual category. To train the model better, a cross-entropy cost function is used as the objective function to measure the difference between y distribution and y^ distribution. The back-propagation mechanism is used to train and update the parameters in the model.

3.5 Credibility Filtering Algorithm

For a given movie review, the key viewpoint sentence is a short sentence that represents the main point of the audience. It is a key part of the commentary emotional information. Through the statistical analysis of the sentiment orientation of key opinion sentences, the fine-grained sentiment analysis is conducted on the whole film reviews. At the same time, the credibility of user reviews is produced.

Splitting an original review by symbols,‘,’、‘。’、‘;’、‘!’、‘?’, into a set of n short sentence units Text Set = {$Text_1$, $Text_2$,…,$Text_m$}; $Review_{orientation}$ is the emotional tendency of the comment. Opi Set = {Opi_1, Opi_2,…,Opi_n} is a set of viewpoint sentences; Ori Set = {Ori_1 Ori_2,…,Ori_n} is a set of viewpoint polarity. Both of them are manually labelled; Define the collection OpinionSet to store all the key information of the viewpoint sentence: Opinion Set = {$opinion_1$, $opinion_2$,…,$opinion_n$}. Each opinion sentence $opinion_i$ consists of unequal short sentence units $Text_j$. $opinion_i$ records the key information of each viewpoint sentence. It includes 3 attributes: $opinion_i$ = {$begin_i$, end_i, $orientation_i$}. $begin_i$ indicates the start position of the short sentence unit of the i-th viewpoint sentence, and end_i indicates the end position of the ith viewpoint sentence. $orientation_i$ indicates the sentimental tendency of the ith opinion sentence, with a value range of {− 1,0,1}. The sentimental tendency of the viewpoint sentence is set to +1 for the positive, 0 for the neutral, and −1 for the negative.

The specific steps of the review credibility calculation method based on the key viewpoint sentence are as follows (Algorithm 1).

Algorithm 1: Method for calculating review credibility based on key opinion sentences

Input: a user movie review, Review=$Text_1, Text_2, \ldots, Text_n$;
Output: The credibility of this film review, $Confidence_1$, $Confidence_2$, $Confidence_3$
Begin
1. for each piece $Opi_i \epsilon OpiSet$, determine
2. if Opi_i is an indivisible unit clause
3. if Opi_i is inseparable $f(i) = 0$;
4. else
$f(i) =$ number of unit sentences by segmenting $- 1$;
5. end if
6. end for
7. The position of the clause: pointer $\leftarrow 1$. Set the view collection OpinionSet to null
8. for i $\leftarrow 1$ to n do;
9. for k $\leftarrow$ pointer to m do ;
10. if $Text_k$ can be found in Opi_i then
11. $begin_i$=k; end_i=k+ $f(i)$; $orientation_i$=Ori_i
12. $opinion_i$={ $begin_i, end_i$, $orientation_i$}
13. Add $opinion_i$ to OpinionSet
14. pointer $\leftarrow$ k + $f(i)$;i = i+1;
15. end if
16 end for
17 end for
18. Confidence $_1$=$Review_{orientation}$ *$\sum_{i=1}^{n} orientation_i$;
19. $Confidence_2$ = $Review_{orientation}$ * $\sum_{i=1}^{n} orientation_i$ * LengthWeigh(i);
20. $Confidence_3$=$Review_{orientation}$ * $\sum_{i=1}^{n} orientation_i$ * Position$Weigh$(i) * LengthWeigh(i);
End

LengthWeigh(i) represents the length of the sentence of the ith opinion sentence $opinion_i$LengthWeigh(i)=$end_i - begin_i$. Position$Weigh$(i) is a normalized quadratic function that gives different weights to viewpoints at different positions:

$$\text{Position}Weigh(i) = 2 * [(\frac{begin_i}{m})^2 - \frac{1}{2}] + 1$$

3.6 Construction of the Overall Emotional Classification Model

The overall emotional classification experiment uses the film reviews with positive credibility as experimental data. It is designed to verify the validity of the credibility algorithm.

The overall sentiment classification model is similar to the key sentence sentiment classification model in Sect. 3.2, except that the Bi-GRU model is replaced by the Bi-LSTM model. Because the data for sentiment analysis at the text layer is a complete review, the length of the sentence is longer than the key sentence. The model of subsequent classification layer is Softmax.

4 Experiments and Results Analysis

4.1 Experimental Parameters and Evaluation

In the experiments, 0.9G movie review data crawled from Douban Website is used as input, and the word2vec's skip-gram model to train word embedding vectors. Each word vector is 200-dimensional. For unregistered words, a uniform distribution of U (−0.01,0.01) is used to randomly initialize the word vector. Other parameters use the default value of the model, and use all-zero initialization < PAD> to terminate a sentence. During the neural network training period, the word vector is allowed to be fine-tuned. After training, a word vector dictionary containing 149,195 words are obtained. The model is trained by 64 samples in each batch, setting Adam's learning rate to 0.001, the cost function to 0.001, Dropout to 0.5, and the number of Bi-GRU units is 256. The hidden layer of the Bi-GRU model is set to 200. If the accuracy does not exceed the current highest accuracy within 20 rounds, the model will automatically terminate early.

The correct rate, recall rate and F value commonly used in the field of natural language processing are used to evaluate the classification result of key viewpoint sentences and the results of overall sentiment classification.

4.2 Fine-Grained Sentiment Analysis Experiment and Result Analysis Based on Key Sentences

Fine-grained sentiment analysis comparison experiments are set up as follows:

(1) RNN. The word vector processed by Word2vec is used as the input of the RNN model, and the output is predicted by the Softmax layer. This experiment is the benchmark model for subsequent experiments.
(2) Bi-LSTM. The word vector processed by Word2vec is used as the input of the Bi-LSTM model, and the output is predicted by the Softmax layer.
(3) Bi-GRU. The word vector processed by Word2vec is used as the input of the Bi-GRU model, and the output is predicted by the Softmax layer.

The related experimental results are shown as in Table 2 and Fig. 3.
From the experimental results, it can be observed that:

(a) In the case of same input and output, the training effect of the RNN model is worse on the test set than those of the Bi-LSTM and Bi-GRU models. It is because the latter can capture the forward and backward bidirectional semantic dependence relationships.

Table 2. Comparison of fine-grained sentiment analysis results

		RNN	Bi-LSTM	Bi-GRU
Overall	P	0.8878	0.8968	0.9026
	R	0.8881	0.8988	0.9034
	F1	0.8877	0.8974	0.9029
Positive	P	0.9039	0.9242	0.9097
	R	0.886	0.9203	0.9287
	F1	0.8948	0.9223	0.9191
Neural	P	0.7976	0.6964	0.7258
	R	0.7283	0.5821	0.6716
	F1	0.7614	0.6341	0.6977
Negative	P	0.8854	0.8974	0.9186
	R	0.9137	0.9193	0.9099
	F1	0.8994	0.9082	0.9142

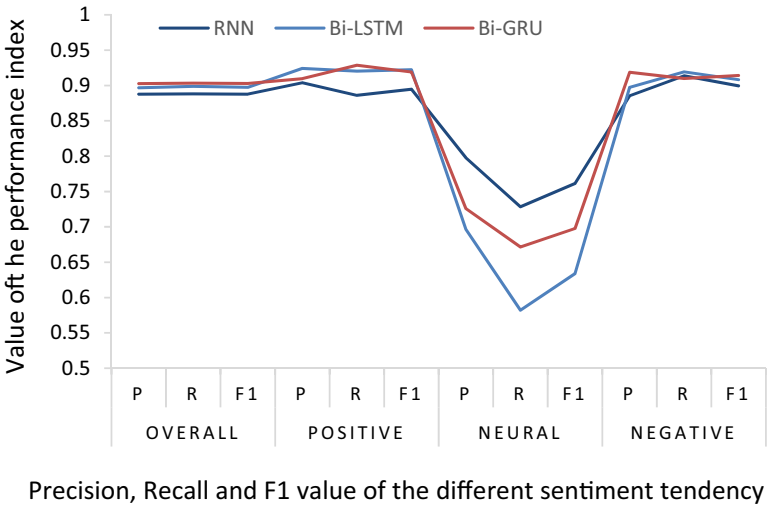

Fig. 3. Graphic illusion of the comparison of fine-grained sentiment analysis results.

(b) For the three basic models in Fig. 3, the emotional prediction results of positive and negative key viewpoints are better than that of the neutral viewpoints because users may have different degrees of deviation when publishing neutral opinions.

(c) From the results in TBALE II, the accuracy and recall rate of the GRU model is higher than that of the LSTM model, and the actual training speed of the GRU model is also faster. It is because the GRU has fewer parameters inside the model than the LSTM model, and the GRU is more suitable for shorter text. Because

of its semantic learning ability, we choose Bi-GRU for the fine-grained sentiment analysis of viewpoint sentences.

4.3 Comment Sentiment Analysis Experiments and Result Analysis Based on Credibility Filtering

This section mainly discusses the impact of each credibility screening on the overall sentiment analysis of film reviews.

The experiments are set up as follows:

(1) Original material. The word vector processed by Word2vec is used as the input of the Bi-LSTM model, and the output is obtained through the Softmax layer. The result of the emotional layer classification of the text layer is obtained.
(2) The first credibility formula (18th line of Algorithm 1) is used to filter the corpus with positive credibility (credibility 1). The word vector processed by Word2vec is used as the input of the Bi-LSTM model, and the output is obtained from the Softmax layer. The effect classification result of the text layer is obtained.
(3) The second credibility formula (19th line of Algorithm 1) is used to filter the corpus with positive credibility (credibility 2). The word vector processed by Word2vec is used as the input of the Bi-LSTM model, and the output is obtained through the Softmax layer. The effect classification result of the text layer is obtained.
(4) The third credibility formula (20th line of Algorithm 1) is used to filter the corpus with positive credibility (credibility 3). The word vector processed by Word2vec is used as the input of the Bi-LSTM model, and the output is obtained through the Softmax layer. The effect classification result of the text layer is obtained.

The obtained experimental results are listed in Table 3, and visualized in Fig. 4.

From the experimental results in TBALE III and Fig. 4, it can be observed that:

(a) In the case of using the original corpus and using the Bi-LSTM and Softmax layers as the base model, the sentiment analysis recall rate and the correct rate of the overall corpus are both low, only at about 70%. It explains that user comments in unfiltered corpus contain high emotional bias.
(b) From Fig. 4, after the credibility filtering of the evaluation, the results of the corpus selected by the credibility 1, 2, and 3 are greatly improved compared with the original corpus. It shows that when users comment the films, there is indeed a situation where the score is inconsistent with the expression emotion, which verifies the validity of our proposed credibility algorithm.
(c) From the comparison, the corpus filtered by credibility 2, 3, the accuracy and recall rate of the sentiment analysis of the text layer are better than the corpus filtered by credibility 1. It shows that the amount of emotional information of the corpus filtered by credibility 2 and 3 is higher, and the actual score of the user is closer to the emotional tendency of the user's comment. The weight of the viewpoint sentence is considered in the calculation of the credibility 2. The location of the opinion sentence and the length-weight of the viewpoint sentence are considered in

Table 3. Experimental results on the different corpus with different credibility filtering

		Origin	Credibility1	Credibility2	Credibility3
Overall	P	0.6983	0.757	0.7845	0.7892
	R	0.6959	0.7551	0.7837	0.7878
	F1	0.6951	0.7553	0.7833	0.7876
Positive	P	0.7259	0.8056	0.7662	0.7578
	R	0.6242	0.7389	0.7516	0.7771
	F1	0.6712	0.7708	0.7588	0.7673
Neutral	P	0.697	0.733	0.8087	0.8212
	R	0.7113	0.7784	0.7629	0.7577
	F1	0.7041	0.755	0.7851	0.7882
Negative	P	0.6688	0.7357	0.7712	0.78
	R	0.7554	0.741	0.8489	0.8417
	F1	0.7095	0.7384	0.8082	0.8097

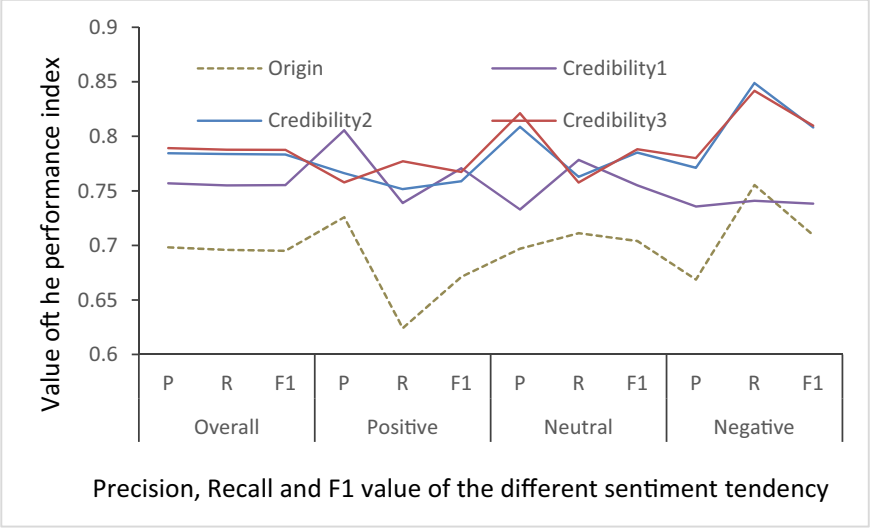

Fig. 4. Graphic illusion of the comparison of user review credibility experiment results.

the algorithm of credibility 3. The result proves that the position and the length of a viewpoint sentence are related to the user's expression of emotion: The longer the length of the viewpoint sentence, the higher the weight of the emotional expression in the text sentence; the viewpoint sentence appearing at the beginning or end of the text has a higher weight than the emotional expression of the viewpoint sentence in a sentence.

5 Conclusions and Future Work

In this paper, a Bi-GRU model is used to perform fine-grained sentiment analysis on key-viewpoint sentences. Based on its result, a film review's credibility algorithm is proposed, where a Bi-LSTM is used to analyze the sentiment of more credible reviews. Corpus screened by the credibility algorithm achieves higher accuracy than the original corpus, which verifies the validity of the credibility filtering algorithm. By considering the fine-grained sentiment analysis tendency, the position of the viewpoint sentence in the text, and the length of the view-point sentence, the credibility produced can be used to filter comments that do not match users rating for sentimental tendency. Furthermore, although the corpus in the experiments is constructed based on the Douban website Dataset, method proposed in this paper is also suitable for the other kinds of film review dataset. The requirement of the dataset is there are sentiment tendencies in the comment sentences and with a grade for the overall sentiment tendency. Therefore, the proposed method has the property of adaptability. However, the situation when users' comments are not related to the movie content is not recognized. Our future research work will focus on the relevance of the evaluation content and the content of the movie, and identify those false user comments.

Acknowledgment. This work is supported by National Natural Science Foundation of China under Grants No. 61671070, 61972364. National Science Key Lab Fund project 6142006190301, National Language Committee of China under Grants ZDI135-53, and Project of Developing University Intension for Improving the Level of Scientific Research–No.2019KYNH226, Qin Xin Talents Cultivation Program, Beijing Information Science & Technology University No. QXTCP B20190.

References

1. Zhao, Y.Y., Qin, B., Liu, T.: Sentiment analysis. J. Softw. **21**(8), 1834–1848 (2010)
2. Chuang, J.L., Xu, S.Z., Hu, Q.F.: Market research report on movie box office, an empirical study of internet word-of-mouth and movie box office. Chin. Bus. Theor. **17**, 6–7 (2016)
3. Huang, S.: Research on fine-grained opinion mining technologies of web review text. Beijing Institute of Technology (2014)
4. Li, Z., Wei, Y., Zhang, Y., et al.: Exploiting coarse-to-fine task transfer for aspect-level sentiment classification. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 33, pp. 4253–4260 (2019)
5. Yang, M., Yin, W., Qu, Q., et al.: Neural attentive network for cross-domain aspect-level sentiment classification. IEEE Transactions on Affective Computing, 1–11 (2019)
6. Pontiki, M., Galanis, D., Pavlopoulos, J., et al.: Semeval-2014 task4: aspect based sentiment analysis. In: Proceedings of the 8th International Workshop on Semantic Evaluation, pp. 27–35 (2014)
7. Rosenthal, S., Farra, N., Nakov, P.: SemEval-2017 task4: sentiment analysis in twitter. In: Proceedings of SemEval (2017)
8. Guo, C., Wang, Z.Y.: Auto-construct of sentiment ontology tree for fine-grained opinion mining. J. Chin. Inform. Process. **27**(5), 75–83 (2014)
9. Liu, L., Wang, Y.H., Hang, W.: Fine-grained sentiment analysis oriented to product comment. J. Comput. Appl. **35**(12), 3481–3486 (2015)

10. Tang, X.B., Lan, Y.T.: Sentiment analysis of microblog product reviews based on feature ontology. Libr. Inform. Serv. **60**(16), 121–127 + 136 (2016)
11. Tang, D.Y., Qin, B., et al.: Effective LSTMs for target-dependent sentiment classification. In: Proceedings of the 26th International Conference on Computational Linguistics, pp. 3298–3307 (2016)
12. Ruder, S., Ghaffari, P., Breslin, J.G.: A hierarchical model of reviews for aspect-based sentiment analysis. In: EMNLP, pp. 999–1005 (2016)
13. Ma, D.H., Li, S.J., Zhang, X.D., Wang, H.F.: Interactive attention networks for aspect-level sentiment classification. In: IJCAI, pp. 4068–4074 (2017)
14. Xue, W., Li, T.: Aspect Based Sentiment Analysis with Gated Convolutional Networks (2018). arXiv preprint, arXiv:1805.07043
15. Guan, P.F., Li, B.A., Lv, X.Q.: Attention enhanced bi-directional LSTM for sentiment analysis. J. Chin. Inform. Process. **33**(02), 105–111 (2019)
16. Zhang, S.Q., Lv, X.Q.: Movie short-text reviews sentiment analysis based on multi-feature fusion. In: International Conference Proceeding Series. ACM (2018)
17. Paul, A.: Who gives a tweet evaluating microblog content value. In: Proceedings of the ACM 2012 Conference on Computer Supported Cooperative, pp. 471–474 (2012)
18. Sheng, Y.: Research on micro-blog information quality evaluation based on content—with sina micro-blog as an example. Inform. Sci. **31**(05), 51–54 + 66 (2013)
19. Hochreiter, S., Schmidhuber, J.: Long short-term memory. Neural Comput. **9**(8), 1735–1780 (1997)