Journal of Corrosion and Materials
www.corromat.com

J. Corros. Mater. **48** (2023) 42–57
https://doi.org/10.61336/jcm2023-5

Sentiment Analysis for E-commerce Product Reviews by Deep Learning Model of Bert-BiGRU-Softmax

Yi Liu[1,*], Jiahuan Lu[1], Jie Yang[1]

[1] *Management School, Hangzhou Dianzi University, Hangzhou, 310018, China*

(Received: 20 June 2023. Received in revised form: 23 October 2023. Accepted: 29 October 2023. Published online: 1 November 2023.)

Abstract

The sentiment analysis has the key problem for the e-commerce product quality management, which customers know other people's attitudes on interested products by e-commerce reviews. Meanwhile, manufacturers are able to learn the public sentiment on their products being sold in E-commerce platforms. This paper proposed the deep learning models of Bert-BiGRU-Softmax, which uses the input layer of Sentiment Bert model to extract multi-dimensional of E-commerce reviews, the hidden layer of Bidirectional GRU model to obtain semantic codes and calculate representing weights of reviews, the Softmax with attention mechanism as the output layer to classify the sentiment tendency. We conduct experiments on a large-scale dataset involving over 500 thousand reviews compared with different learning models. The experimental results show that the proposed models reach the high accuracy 95.5% on the E-commerce reviews, and the outperforms of RNN, BiGRU, Bert-BiLSTM in terms of accuracy and loss.

Keywords: Sentiment Analysis, E-commerce Product Review, Bert, Bidirectional GRU.

1. Introduction

With the rapid growth of e-commerce, online shopping can bring convenience and low price products for consumers. Due to the information inconsistency between the product quality and description provided by the seller, more and more consumers will seek product information from E-commerce reviews involved many aspects of products: appearance, price, logistics and so on. Online E-commerce reviews are very useful for customers in shopping decision process, meanwhile the manufactures and E-commerce platform can improve their products and on-line services by the public opinions at sometimes [1]. Almost all of the E-commerce platform, such as Amazon (http://www.amazon.com) and Taobao (http://www.taobao.com, the biggest online trading website in China), provide a reviewing system enabling customers to share their opinions towards one or more products.

With the enormous amount of E-commerce reviews rapid growth on the Web platform, customers are difficult to obtained the valuable and reliable online shopping auxiliary information. Traditional sentiment analysis was usually performed in terms of a document-based granularity or a sentence-based granularity, which objective is to identify the positive/negative opinion expressed in the whole document or sentence. Pavlou and Dimoka (2006) utilized content analysis to quantify over 10,000 publicly available feedback text comments of 420 sellers in eBay's online auction marketplace, which can be differentiated among sellers and prevent a market of lemon sellers for online market [2]. The sentiment analysis is the text content analysis technique to distinguish the positive or negative attitudes and opinions of consumer from the related information appeared in different forms like BBS, blogs, Wiki or forum websites (Abbasi et al., 2008) [3]. Although statistical learning methods have achieved success in e-commerce platform product review sentiment classification, two problems have limited its practical application: 1) The computational efficiency to process large-scale reviews; 2) the ability to continuously learn from increasing reviews and multiple domains. However, as a piece of E-commerce review generally involves many aspects of product, such as appearance, quality, price, logistics, and so on. In this paper, we define each of these aspects as a dimension of a product. In general, a customer may be satisfied with some dimensions of a product but dislike some other features, which is not suitable to conduct document-level or sentence-level sentiment analysis for E-commerce reviews.

Therefore, the challenges in multi-dimensional sentiment analysis lie in dimensions mapping and sentiment word disambiguation. The dimension mapping problem refers to mapping opinioned text blocks with right dimensions. For instance, the text block "Sounds good!" will be mapped with the dimension "Audio". The sentiment word disambiguation problem refers that a sentiment word may be connected with two or more dimensions. A better and more suitable way is to perform sentiment analysis on separated dimensions of products, which execute a multi-dimensional sentiment analysis on E-commerce reviews. In this paper, we focus on those two problems and proposes the deep learning hybrid model of Bert-BiGRU-Softmax with the attention mechanism to finally extract dimension-oriented opinions from E-commerce reviews,

*Corresponding author (liuyi@hdu.edu.cn)

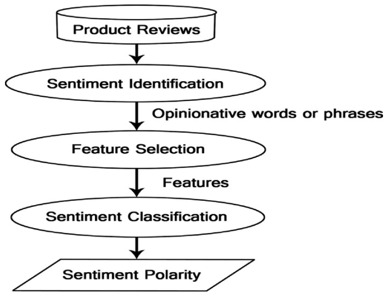

Figure 1: Sentiment analysis process on E-commerce product reviews.

which utilized the Bert to solve the dimension mapping and sentiment word disambiguation problems of e-commerce feature text and the BiGRU with the attention mechanism to remember long-distance dependent information for multi-dimensional sentiment analysis of the e-commerce product quality reviews.

2. Related work

Sentiment Analysis (SA) has always been the hot research topic in recent years, which can be considered a classification process as shown in Figure 1. There are three main classification levels in Sentiment Analysis: document-level, sentence-level, and aspect-level tasks. Document-level SA aims to classify an opinion document as expressing a positive or negative sentiment. Sentence-level SA aims to classify sentiment expressed in each sentence. Aspect-level SA aims to classify the sentiment with respect to the specific aspects of entities. Different approaches have been proposed to realize the goals of those tasks. The first approach is called semantic-based approach (Zhang and Ye, 2010) [4], which performs sentiment analysis based on rules and sentiment polarity lexicons, and the second approach is machine- learning methods, which regards sentiment analysis as a binary or multi-class classification task and uses common classification methods(Chang and Lin CJ Ye, 2011) [5].Most of the previous works on sentiment analysis can be put in the above two types, but the latest sentiment analysis method is the deep learning models including the convolutional neural network (CNN), recurrent neural network (RNN) model and other models (Hu F and Li L, 2017) [6].

2.1 Feature Selection Methods

Sentiment Analysis task is considered a sentiment classification problem. The first step in the SC problem is to extract and select text features. Some of the current features are [7]:

Terms presence and frequency: These features are individual words or word n-grams and their frequency counts. It either gives the words binary weighting (zero if the word appears, or one if otherwise) or uses term frequency weights to indicate the relative importance of features [8]. Parts of speech (POS): finding adjectives, as they are important indicators of opinions.

Opinion words and phrases: these are words commonly used to express opinions including good or bad, like or hate. On the other hand, some phrases express opinions without using opinion words. For example: cost me an arm and a leg.

Negations: the appearance of negative words may change the opinion orientation like not good is equivalent to bad.

Feature selection methods attempt to reduce the dimensionality of the data by picking from the original set of attributes, which can be divided into lexicon-based methods that need human annotation, and statistical methods that are more frequently used. Lexicon-based approaches usually begin with a small set of 'seed' words. Then they bootstrap this set through synonym detection or on-line resources to obtain a larger lexicon. This proved to have many difficulties as reported by Whitelaw et al [9]. The mutual information measure provides a formal way to model the mutual information between the features and the classes. Turney and Littman (2003) used the point wise mutual information (PMI) to extend the commendatory and derogatory emotional vocabulary in order to analyses the emotional tendency of English text. Yu and Wu [10] have extended the basic PMI by developing a contextual entropy model to expand a set of seed words generated from a small corpus of stock market news articles. Their method outperformed the (PMI)-based expansion methods as they consider both co-occurrence strength and contextual distribution, thus acquiring more useful emotion words and fewer noisy words.

Feature transformation methods create a smaller set of features as a function of the original set of features. Latent Semantic Indexing (LSI) is one of the famous feature transformation methods. LSI method transforms the text space to a new axis system which is a linear combination of the original word features. Principal Component Analysis techniques (PCA) are used to achieve this goal [11]. There are other statistical approaches which could be used in FS like Hidden

Markov Model (HMM) and Latent Dirichlet Allocation (LDA). They were used by Duric and Song [12] to separate the entities in a review document from the subjective expressions that describe those entities in terms of polarities. HMM-LDA is a topic model that simultaneously models topics and syntactic structures in a collection of documents [13].

2.2 Sentiment Classification techniques

Sentiment classification techniques can be divided into machine learning approach, lexicon based approach and hybrid approach [14]. The Machine Learning Approach (ML) applies the famous ML algorithms and uses linguistic features. The Lexicon-based Approach relies on a sentiment lexicon, a collection of known and precompiled sentiment terms. It is divided into dictionary-based approach and corpus-based approach which use statistical or semantic methods to find sentiment polarity. The hybrid Approach combines both approaches and is very common with sentiment lexicons playing a key role in the majority of methods.

Kim and Hovy (2004) analyzed the emotional tendency of text based on the synonyms, antonyms and hierarchies of the WordNet dictionary [15]. The acquisition of sentiment tendency about Chinese text mainly depended on HowNet [16]. Y L Zhu and J Min (2006) used the semantic similarity and correlation of HowNet to calculate the similarity between the new word and the datum word, thereby to distinguish the sentiment tendency of the text [17]. G. Somprasertsri (2006) mining product feature and opinion of online customer reviews based on the consideration of syntactic information and semantic information[18]. Sentiment analysis model of sentiment polarity lexicons is commonly used, but the existing sentiment vocabulary lexicons is limited (B Liu,2014) [19].

Compared with sentiment polarity lexicon, Machine learning methods have better advantages in the nonlinear and high dimensional pattern recognition problem. Pang, et al (2002) firstly applied the machine learning method of N-Gram into the emotional analysis field, the experiment result show the N-Gram reached highest classification accuracy of 81.9% [21]. Go A. and Bhayani R (2009) proposed the supervised learning method to classify the positive or negative emotional reviews extracted from Twitter by the naive Bayes, maximum entropy and SVM analysis algorithms, which the experiment show 80% analysis accuracy [22]. Marco Guerini (2013) apply the sentiment analysis methods such as Naive Bayesian (NB), K-Nearest neighbor (kNN), Maximum entropy(ME), support vector machine (SVM) to analyses the reviews emotional tendency, which the support vector machine (SVM) is obviously superior to other methods in large training set and reach the highest accuracy of 83% [23]. Due to the feature selection will affect the performance of the machine learning method, Abinash Tripathy (2016) analyzed online comment reviews by the N-gram model combined machine learning methods, experiments show that the SVM combined unigram, bigram and trigram features achieved the best classification results [24].

With the development of deep learning research, the deep neural network presents outstanding performance in natural language processing. Kim (2014) proposed using convolutional neural network (CNN) to solve sentiment classification problem and achieved good results. Santos used deep convolution neural network to analyze sentiment in short texts[25]. The use of advanced architectures of RNNs, Cho et al. (2014) proposed gated recurrent unit (GRU) for learning long dependencies has led to signi?cant improvements in various tasks[26]. Qu Zhaowei and Wang Yuan (2018) proposed a sentiment analysis model based on hierarchical attention network, which improved by five percentage points compared with the traditional recurrent neural network [27]. Trofimovich J. (2016) used recurrent neural networks (RNN) to model sentences, and LSTM (long short-term memory), which is a model of recurrent neural networks, was also proved effective to solve sentiment analysis problems[28].Despite the effectiveness and simplicity of this approach, the results wrought from the traditional approach are far from optimal and call for novel introduction of different methods for improved accuracy outside of just changing the different types of word embedding (i.e. GloVe or BERT) [29]. This is caused by the fact that the word embedding may not be able to completely capture the entire context of a sentence or review especially as the review size becomes larger (Devlin et al., 2019)[30]. Moreover, variants of RNNs, such as BiLSTM (Bidirectional-LSTM), BiGRU (Bidirectional-GRU), are exploited for text classification, specifically in sentiment analysis as exhibited in [31]. In order to manage the sentiment by focusing on phrase-level classification containing linguistics such as negativity, intensity, and polarity, a model centered on the regularization of LSTM enlightened by [32]. To handle syntactic and semantics of the text, a Japanese language network model by using bidirectional LSTM with the training of tagging offered in [33]. In polarity classification, a hierarchical multi-input and output model (HMIO) for phrase-level demonstration use independents BD-GRU (Bi-Directional Gated Recurrent Unit) proposed in [34].

There are numerous techniques; however, we are intended by work effectively done on joint neural networks with unsupervised learning [35]. In this paper researched the Bert-BiGRU-Softmax models for sentiment analysis of the e-Commerce product quality reviews, which the sentiment Bert model make the features exaction by mapping each sentences into the proper dimension respectively, Bi-directional GRU model solve the sentiment word disambiguation problems by using the multiple clause-rules recognition, and the Softmax model with attention mechanism make the multi-dimensional

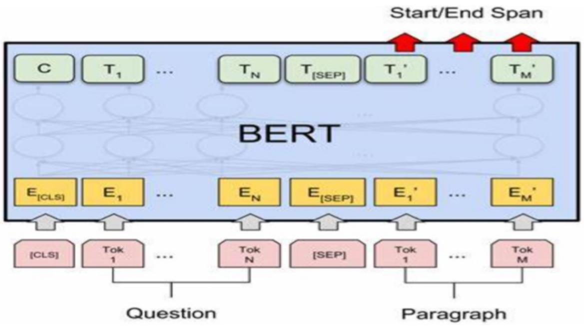

Figure 2: The multi-level structure chart of Bert Model

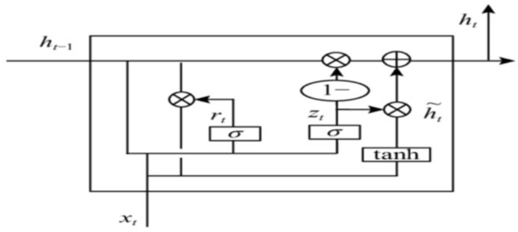

Figure 3: Information structure of GRU Model

sentiment analysis by calculated the sentiment polarity of specific dimension of the e-Commerce product reviews.

3. Deep-Learning Models

3.1 BERT Model

The BERT (bidirectional encoder representations from transformers) is the transfer learning-pre-training neural network model, which is recent paper published by Google AI researchers at 2018. The BERT apply the bidirectional training of transformer to learns contextual relations between words in text, which transformer includes two separate mechanisms that an encoder that reads the text input and a decoder that produces a prediction for the task. As opposed to single-direction models of RNN (Recurrent Neural Network) and LSTM (Long Short-Term Memory), which read the text input sequentially (left-to-right or right-to-left), the transformer encoder reads the entire bidirectional sequence of words at once. It has caused a stir in the machine learning community by presenting state-of-the-art results in a wide variety of NLP (natural language processing) tasks and others. The figure 2 is the multi-level structure chart of Bert model.

The novel technique of BERT is the masked language model (MLM) which allows bidirectional training in models. This characteristic allows the BERT model have deeper sense of language context and learn the context of word by all of its surrounding, which it was previously impossible.

3.2 GRU Model

The GRU (Gated Recurrent Unit) is part of a specific model of recurrent neural network that intends to use connections through a sequence of nodes to perform machine learning tasks associated with memory and clustering, which allows GRU to carry forward information over many time periods in order to influence a future time period. GRU can be considered as a variation on the LSTM because both are designed similarly and produce equally excellent results, which gated recurrent units help to adjust neural network input weights to solve the vanishing gradient problem that is a common issue with recurrent neural networks [36] . The structure chart of the GRU is shown in the Figure 3.

As a refinement of the general recurrent neural network structure, GRU have what's called an update gate zt and a reset gate rt. Using these input vectors xt and output vectors ht, the model refines outputs by controlling the ht?1 flow of information through the model. Like other kinds of recurrent network models, GRU with gated recurrent units can retain information over a period of time t that is why the simplest ways to describe these types of technologies as the "memory-centered" type of neural network. By contrast, other types of neural networks without gated recurrent units often do not have the ability to retain information.

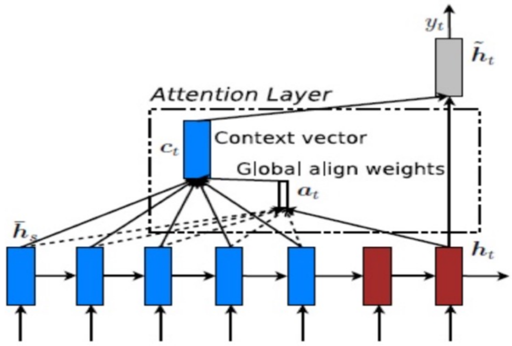

Figure 4: Information structure of Attention Mechanism

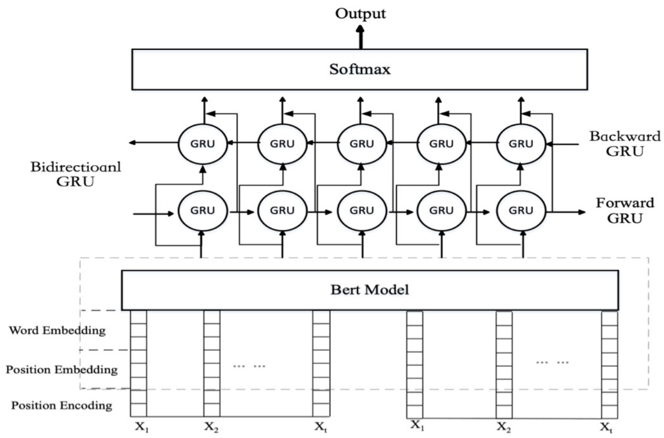

Figure 5: Information flow chart of the Bert-BiGRU-Softmax Model

3.3 Attention Mechanism

Before Attention mechanism, translation relies on reading a complete sentence and compress all information into a fixed-length vector, which the sentence with hundreds of words represented by several words will surely lead to information loss and inadequate translation, etc. attention partially fixes this problem. The attention mechanism proposed by Bahdanau, et al. (2014) to help memorize long source sentences in neural machine translation (NMT) and NLP[37]. Attention probability distribution matrix of the input text sequence is obtained by attention mechanism. The weight of text feature information is calculated by the matrix, which reduces the missing and redundant information during feature extracting. It allows machine translator to look over all the information the original sentence holds, then generate the proper word according to current word it works on and the context, the structure of the attention mechanism is shown in the Figure 4.

4. Sentiment analysis of e-Commerce product review by Bert-BiGRU-Softmax model

This paper proposed the hybrid models of Bert-BiGRU-Softmax for sentiment analysis on the e-commerce product quality reviews. The sentiment Bert model, as input layer, make the feature extraction at pre-processing-phase. Further, the hidden layer of Bidirectional GRU perform dimension-oriented sentiment classification by using bidirectional long short-term memory and gated recurrent Units to hold the long-term dependencies, which are inherent in the text regardless of lengths and occurrences. While in the post-processing-phase, the output layer of Softmax calculate the sentiment polarity by comprised of pooling to smaller dimensions by weighted attention mechanism. The Figure 5 shows the proposed models structure.

4.1 Sentiment-Bert for Dimension Mapping

Dimensions mapping of the e-commerce product quality reviews are typically domain-dependent problem. Since a dimension may appear in many product categories, we have to determine the right dimension for a given review about a certain e-Commerce product category.

In this study, The Bert transform the e-Commerce product reviews into a numerical matrix where each column represents the identified features and each row represents a particular review. The matrix is given as input to Bert algorithm in order to train the model by two training strategies of the MLM and next sentence prediction, with the best performance of the

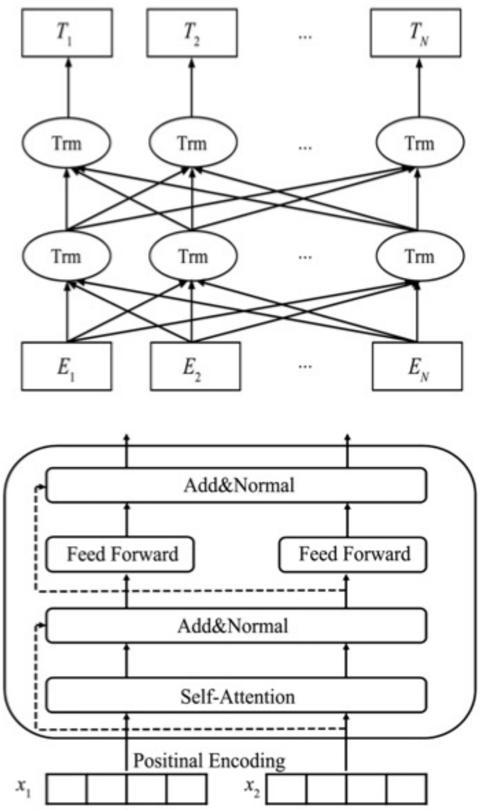

Figure 6: The pre-trained model and transformer coding unit of Bert model

combined loss function[38] . When training the BERT model, sentences segmentation procedure is used for splitting a long sentence into several shot text blocks, and word sequences are regarded as nodes of text blocks, which is more efficiently and easier to map those short sentences into corresponding dimensions. The transformer receives pairs of sentences as input and learns to predict the second sentence in the pair on the assumption that the random sentence will be disconnected from the first sentence, which can be taken as explicit features in dimension mapping. The training process of Bert model is shown in the Figure 6.

Aiming at sentiment task and complex emotional features, the pre-training approaches of the sentiment Bert model is focusing on learning sentiment features by context-based word sentiment prediction task, which classify the sentiment of masked words to acquire the textual representation biased towards sentiment features.

The sentiment Bert model can be used effectively to learn feature extraction over a variable-length sequence S by learning the distribution over the next input text vector. Given a review sentence S, we can obtain its category c and the dimension set Dc of category c directly. For each word $w_i(w_1, w_2, w_3, ..., w_m)$ in the review sentence S and a dimension $d_j \in D_c$, we assign a probability score which describe the probability $p(S)$ of word w_i belongs to category d_j in the e–Commerce product quality reviews (as shown in Formulas 1).

$$\mathrm{p}\left(\mathrm{S}\right) = \mathrm{p}\left(\mathrm{w}_1, \mathrm{w}_2, \ldots, \mathrm{w}_\mathrm{m}\right) = \prod_{\mathrm{i}=1}^{\mathrm{m}} \mathrm{p}(\mathrm{w}_\mathrm{i}|\mathrm{w}_1, \mathrm{w}_2, \ldots, \mathrm{w}_{\mathrm{i}-1}). \tag{1}$$

Transformer add sequence information by the position embedding formulas (2) and (3).

$$\mathrm{P}\left(\mathrm{pos}, 2\mathrm{i}\right) = \sin\left(\mathrm{pos}|10000^{\frac{2i}{d_{model}}}\right) \tag{2}$$

$$\mathrm{P}\left(\mathrm{pos}, 2\mathrm{i}{+}1\right) = \cos\left(\mathrm{pos}|10000^{\frac{2i}{d_{model}}}\right) \tag{3}$$

Where d is 64 , the text sequence be represented as the 512 characters, $2i$ is even position and $2i+1$ is odd position in the given sequence of input vector.

When the Transformer make the feature extraction as $\mathrm{x}_{[CLS]}$ and $\mathrm{x}_{[MASK]}$ from the special words of $\mathrm{w}_{[CLS]}$ and $\mathrm{w}_{[MASK]}$ in the S sequence, the Bert loss function takes into consideration only the prediction of the masked values and ignores the prediction of the non-masked words, which be represented as formulas (4-5).

$$loss = \sum_i L(x^i_{[MASK]}) F(x^i_{[MASK]}) \tag{4}$$

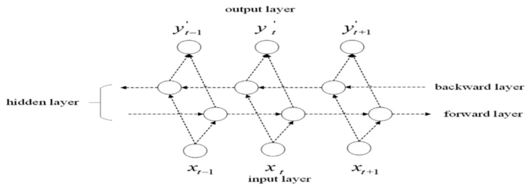

Figure 7: The structural chart of the bidirectional GRU network.

$$F\left(x^i_{[MASK]}\right) = \begin{cases} k \;\; if \; x \in R \\ 1 \; if \; x \notin R \end{cases} \tag{5}$$

where R is the emotional words lexicon of e-Commerce product quality, $L(x^i_{[MASK]})$ is the loss of the masked words $x^i_{[MASK]}$ in prediction process of input sequence sentence, $F(x^i_{[MASK]})$ is the weight of emotional and non-emotional words. The sentiment Bert model involves the following steps:

Sentiment Bert model:
Begin
In this step, each target sentence of the e-Commerce product quality review is represented by the content of the sentence. Before feeding word sequences into BERT, 15% of the words in each sequence are replaced with a [MASK] token. The model attempts to predict the original value of the masked words, based on the context provided by the other, non-masked, words in the sequence. We use this representation to learn the internal features of the sentence.
Sentence contextual information processing:
Here, we adopt the method that simply replace each word of the target sentence with special characters to build the meaningless string of [MASK], which is replaced by a [MASK] string to obtain the contextual information of each sentence. For example, "The mobile phone was [MASK] and the service was good", the unmasked description of "the service was good" is the positive assessment of mobile phone,conjunction "and" express the similar sentences, so the Masked words can be deduced as the positive expression.
Sentiment Bert integrate processing:
This step integrates the content and contextual information of the sentence to construct the masked sentence model. It implements the integration by inputting the above two training samples together to the Bert fine–tuning procedure to train the masked sentence model for recognition task. For each dimension, we put the top 20 words with highest scores into the candidate sentiment word set. As consequence, Bert model increase the sentiment analysis accuracy of reviews by bidirectional context awareness.
End

4.2 Bidirectional GRU for Dimensional Sentiment

Sentiment lexicons consist of the sentiment polarity in e-commerce product quality reviews and are widely used in sentiment analysis. However, those lexicons did not distinguish the polarities of one word on different dimensions about product. We use multiple rules to improve the performance of sentiment analysis. For example, "It is so famous brand of mobile phone that the customer like to buy Huawei." is recognized as a positive strong rule, which infers that sentences matching with this rule will be labelled as positive without considering other factors. The emotion analysis output have the relationship with the previous state and the next state of the text sequence at same time. the GRU is one-way neural network structure of this fixed-length context vector design, which it has forgotten the first part once it completes processing the whole input.

This paper proposed the bidirectional gated recurrent unit (BiGRU) to perform natural language recognition tasks associated with memory and clustering by both forward and reverse information (Grü.er-Sinopoli and Thalemann [39]). The idea of BiGRU is to split the regular GRU neurons into a forward states (positive time direction) and a backward states (negative time direction). Based on the co-occurrence information with seed sentiment words, we could enlarge the original seed words iteratively. Besides, BiGRU connects two hidden layers with opposite transmission directions to the same output layer, so that the output layer obtains information both from past and future states. Therefor the BiGRU is able to learn the information from two different data directions, and make more accurate analysis of the e-commerce product quality reviews based on dimensional lexicon and rules. The general structure of BiGRU is shown in Figure 7.

BiGRU have the given sequence of input vectors $<x_1,x_2,\ldots, x_t>$ (x_t represents the concatenations of input features), setting unit computes the corresponding hidden activations $< h_1, h_2,\ldots,h_t >$, and outputs a vector sequence $<y_1,y_2,\ldots,y_t$

$>$ from the input data. At time t, the current hidden state of h_t is determined by three parts including input vectors $<x_1, x_2, \ldots, x_t>$, the forward hidden state $\overrightarrow{h_t}$ and the backward hidden state $\overleftarrow{h_t}$ at the time. The reset gate (r_t) controls the extent to which status information from previous time is ignored; the smaller the r_t, the more previous status information is ignored. The update gate (z_t) controls the degree to which new input information is received in the cell state. The symbol $\otimes$ is the element–wise multiplication. σ is the sigmoid function, and tanh is the tanh function. The hidden state h_t, update gate (z_t) and the reset gate (r_t) of BiGRU are calculated as shown in Eqs. (6)–(10).

$$z_t = \sigma\left(w_z . \left[h_{t-1}, x_t\right]\right) \tag{6}$$

$$r_t = \sigma\left(w_r . \left[h_{t-1}, x_t\right]\right) \tag{7}$$

$$\overrightarrow{h_t} = GRU(x_t, \overrightarrow{h_{t-1}}) \tag{8}$$

$$\overleftarrow{h_t} = GRU\left(x_t, \overleftarrow{h_{t-1}}\right) \tag{9}$$

$$h_t = w_t \times \overrightarrow{h_t} + v_t \times \overleftarrow{h_t} + b_t \tag{10}$$

where the GRU( ) is the nonlinear transformation function for the input vectors. w_z and w_r are the weight matrices of update gate and reset gate. v_t and w_t are the weight matrices of the forward and backward hidden state. b_t is the corresponding offset of hidden state at time t. $*$ denotes the element-wise multiplication, and [] indicates that two vectors are connected to each other.

The approaches of BiGRU model involve the following steps:

BiGRU Model:

Begin

Update gate: This gate calculate the update gate Z_t at time step t executing the following steps:

(1) The input X_t is multiplied by a weight W_{zx};

(2) The previous output H_{t-1} which hold information from previous units multiplied by weight W_{zh} ;

(3) Both are added together and sigmoid function apply to squeeze the output between 0 and 1.

Reset gate: This gate calculate the update gate R_t at time step t executing the following steps:

(1) The input X_t is multiplied by a weight W_{xr};

(2) The forward output $\overrightarrow{h_{t-1}}$ which hold information from forward units multiplied by a weight W_{hr};

(3) Both are added together and sigmoid function apply to squeeze the output between 0 and 1.

This gate has the opposite functionality in comparison with the update gate since it is used by the model to decide how much of the past information to forget.

Current memory content: The calculation involves the following steps:

(1) The input X_t is multiplied by a weight W_x

(2) Apply element-wise multiplication to the reset gate R_t and the backward output layer $\overleftarrow{h_{t-1}}$; this allows to pass only the relevant past information.

(3) Both are added together and a tanh function is applied.

Final memory at current time step: Last but not least the unit has to calculate the H_t vector which holds information for the current unit and it will pass it further down to the network. Key role in this process plays the update gate Z_t.

(1) Apply element-wise multiplication to the update gate Z_t and H_t.

(2) Apply element-wise multiplication to one minus the update gate Z_t and H_t.

(3) Both are added together.

End

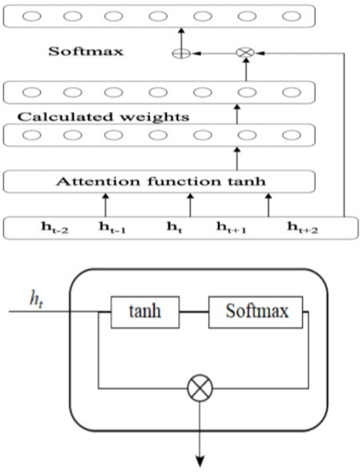

Figure 8: The information flow chat of Softmax function with attention mechanism

Table 1: The experiment environment configuration

Environment	Configuration
Windows Platform	Win 8.1
Python	Python 3.5
Deep Learning Framework	Keras & Tensorflow
Software	JetBrains PyCharm

4.3 Softmax with Attention mechanism for Sentiment Analysis

The output layer of the Softmax function will output the positive scores, negative scores, and neutral scores by fusing the different semantic features of the Bert-BiGRU models, the attention mechanism calculate the linear weighted sum of all the positive, negative, and neutral polarities of the sentence sequences from last hidden state level [40], which the attention probability and sentiment polarity weights of reviews represented are calculated by the Softmax function (as shown in Figure 8).

The detailed approaches of the Softmax function with attention mechanism are as following:

1) After received the feature vector of attention layer, Make the attention function of weight calculation matrices w_a:

$$u_t = \tanh(w_a h_t + b_a) \tag{11}$$

$$\alpha_t = \frac{\exp(u\,{}_t^T u_a)}{\sum_t \exp(u\,{}_t^T u_a)} \tag{12}$$

$$s_{it} = \sum_{i=1}^{n} a_{it} h_{it} \tag{13}$$

where tanh is the sigmoid function, the hidden state h_t is fed into learning function (11) to produce probability vector u_t, the vector α_t is computed as averaged weight.

2) Make the normalization operation by Softmax function (14):

$$y = softmax(w_a s_{it} + b_a) \tag{14}$$

where y is the sentiment analysis results, b_a is the corresponding offset of output layer.

5. Experimental Design and Result Analysis

5.1 The experiment environment

For experiments involving Bert-BiGRU, we employed the deep learning framework of the Keras and Tensorflow, experimental environment on Win8.1 system, Python 3.5, JetBrains PyCharm. The running configuration of deep learning model are as shown in Table 1:

The performance evaluation indicators of deep leaning models are accuracy, loss, F1 measure and so on, which are calculated by the confusion matrix of the analysis problem, as shown in Table 2.

Table 2: Evaluation indicators of confusion matrix

Real situation	Prediction results	
	Positive	Negative
positive	True positive(TP)	False negative(FN)
negative	False positive(FP)	True negative(TN)

Table 3: The performance of different deep learning model

Model	Accuracy	F1
Multi-Bi-LSTM[44]	0.8840	0.8840
Bert-BiGRU	0.8903	0.8857
WWAL(Weight W2V-ATT-LSTM)[45]	0.8727	0.8719
Bert-BiGRU	0.8831	0.8816

The accuracy is the percentage that correct predicted samples to the total number of all the test sets, which is calculated by formula (15). F1 measure combines recall and precision in the formula (16). We consider Precision, Recall, F-Measure, and Receiver Operating Characteristic (RoC) as the performance metrics.

$$\text{Accuracy} = \frac{(\text{TP+TN})}{(\text{TP+FN+FP+TN})} \tag{15}$$

$$\text{F1} = \frac{(2 \times \text{Recall} \times \text{Precision})}{((\text{Recall+Precision}))} \tag{16}$$

5.2 Experiments and Results Analysis

5.2.1 Experiment 1

To precisely assess the applicability, efficiency, and reliability of the Bert-BiGRU Model, we have chosen following datasets of multi-source corpora with various domains and size respectively. The first dataset is IMDB [41], which is used and prolonged extensively as a benchmark dataset [42]. The dataset contains 25 000 tweets with the polarity of 12,500 positives, and 12,500 negatives about of the movie reviews. The second dataset is the famous Chinese emotional corpus known as ChnSentiCorp [43], which has abundant Sentiment Corpus including ChnSentiCorpHtl ChnSentiCorpMov, and so on. To procure Chinese opinion analysis evaluation, we make experiments on ChnSentiCorp-Htl-ba-6000 and COAE2014-task4 by proposed emotion analysis model compared with the multi-Bi-LSTM model [44]and the Weight W2V-ATT-LSTM (WWAL) model [45]. Tables 3 report the performances of different models.

From tables 3 we can draw following conclusions that Multi-Bi-LSTM, Weight W2V-ATT-LSTM and Bert-BiGRU are all suitable for sentiment analysis. But the Bert-BiGRU shows better performance in review sentiment analysis by Bert model as pre-training text feature extractor. In first observation of COAE2014-task4, Accuracy and F1 measure of Bert-BiGRU is (0.8903, 0.8857), which is one percent larger than Multi-Bi-LSTM (0.8840, 0.8840). The second observation of ChnSentiCorp-Htl-ba-6000, Bert-BiGRU produces the better accuracy (0.8831) and F1 measure (0.8816) which is slightly higher than accuracy (0.8727) and F1 measure (0.8719) of WWAL. When the feature classifiers are transferred across different topics, the observation results inspire us a rule: the machine learning techniques for sentiment classifier is severely dependent on its domains.

5.2.2 Experiment 2

This paper crawled and analyzed a large-scale dataset with 150 predefined dimensions of 500 thousand E-commerce reviews about mobile phone products from Sunning, Taobao and other e-commerce websites, which covers almost all the aspects of different product with their polarity concerning positive and negative (e.g., the dimensions ”quality”, ”logistic” and ”service”). we use the deep learning models of RNN, BiGRU, Bert-BiLSTM and Bert-BiGRU to make sentiment analysis of E-commerce product quality reviews from dimensions of brand, ratings, price and others. Table 4 shows the dataset about the mobile phone quality reviews. ”trust”, ”wonderful” and ”good” have top-most scores. ”poorly”, ”frustrated” and ”unsatisfied” have lowest scores. The high scores for ”wonderful” and ”good” could be the newly delivered phones. Also, the highest score for ”trust” among all the emotions shows that the reviewers are conviction and trust the product.

This paper compares the performances of RNN, BiGRU, Bert-BiLSTM and Bert-BiGRU models according to the indicators of accuracy and loss value, which are based on the 400 thousand original data from the web reviews corpus are selected as training sets, and the 100 thousand original data as test sets in experiment. The hyper parameters setting of Bert-BiGRU model are as follow:

Table 4: The sentiment reviews about mobile phone quality on e-Commerce websites

1	This mobile phone provided bigger screen and powerful battery,the only drawback for me is that it's a little thick ...
2	This mobile phone isn't famous brand, but price is cheaper than others and the battery life is about 2300 mAh. The cost performance is very high!
3	The mobile phone has wonderful software function and appearance which could be better. A friend recommended it to me..., is Although charge
4	I admit, mobile phone has good camera system with high pixel. If anything, it helps you make beautiful picture what you want .
5	Poor battery efficiency and signal, I have to charge this mobile phone so many times in a day! It's so turgid that I put it down in frustration. ...
6	The particularly nice mobile phone is received, the audio sounds good! courier delivery it to my home, give five star rating for it ...

Table 5: The hyper parameters setting of Bert-BiGRU model

Hyper parameter name	Parameter value
The nodes of hidden layer network	100
Loss function	Categorical_Crossentropy
Optimizer	Adam
Dimensions of word vector	768
Epoch	10

Table 6: The accuracy value of different deep learning models.

Model	Running Times				
	1	2	3	4	5
RNN	0.867	0.876	0.895	0.893	0.886
BiGRU	0.890	0.910	0.918	0.913	0.921
Bert-BiLSTM	0.881	0.906	0.927	0.922	0.911
Bert-BiGRU	0.912	0.940	0.934	0.939	0.929
Model	Running Times				
	6	7	8	9	10
RNN	0.894	0.893	0.877	0.894	0.893
BiGRU	0.921	0.921	0.922	0.921	0.922
Bert-BiLSTM	0.919	0.928	0.927	0.931	0.925
Bert-BiGRU	0.944	0.936	0.955	0.952	0.947

Table 7: The loss value of different deep learning model.

Model	Running Times				
	1	2	3	4	5
RNN	0.342	0.329	0.303	0.300	0.346
BiGRU	0.287	0.235	0.240	0.362	0.257
Bert-BiLSTM	0.349	0.291	0.223	0.356	0.389
Bert-BiGRU	0.262	0.192	0.234	0.239	0.325
Model	Running Times				
	6	7	8	9	10
RNN	0.359	0.378	0.427	0.410	0.428
BiGRU	0.307	0.366	0.374	0.439	0.427
Bert-BiLSTM	0.343	0.343	0.395	0.412	0.385
Bert-BiGRU	0.269	0.235	0.211	0.291	0.252

The accuracy and loss value of different deep learning models on the dataset are shown in Table 6 and Table 7. The accuracy and loss Cure of different models on the dataset are shown on Figure 9 and Figure 10.

According to the accuracy cure, the RNN model reach the highest accuracy value of 0.895 at the 3rd run-time, and then the accuracy value of RNN became volatility at the over-fitting status. The BiGRU model reach the highest accuracy value of 0.922 at the 8th run-time, and then the accuracy value ascend to the over-fitting status. The Bert-BiLSTM model reach the highest accuracy value of 0.931 at the 9nd run-time, and then the accuracy value of LSTM became descend to the over-fitting status. dimension reduction. Compared with the original model BiGRU, the proposed Bert-BiGRU model reach the highest accuracy value of 0.955 at the 8th run-time and reduce computational overhead, which is superior to the competitor algorithms.

According to the loss value cure of different models, the extraction ability of RNN is the worst of all models. The BiGRU model will generate large computational overhead and poor accuracy performance. The Bert-BiLSTM model can reach the higher performance than that of the BiGRU. The proposed Bert-BiGRU model reach the highest accuracy value of 0.955, which is improved by nearly 3.6% points of Bert-BiLSTM at the training datasets. By compared the models from different types of comment sets, experiment show that the Bert-BiGRU model can change dimension flexibly and improves the accuracy of feature extraction, which is applied for the sentiment analysis of e-commerce product quality reviews.

Furthermore, this paper uses the sentiment Bert-BiGRU- Softmax model to extract the quality features (camera, screen, system, audio, battery, signal, service, price and surface), obtain each type polarities including the attention probability distribution from input text sequences, calculate the weights of customer reviews emotional scores according to the different dimensions of mobile phone products. The sentiment analysis results (positive, neutral and negative) of e-commerce product quality reviews are shown in Table 8.

After the comprehensive compared, AP mobile phone get higher satisfaction at the features of system, hardware and service than the other three brands form consumer, as while the satisfaction of signal and price are the lowest of the four brand products, which can speculate the consumer should pay attention to the signal instability problem. Mi mobile phone has the highest satisfaction form consumer at the feature of price, surface and audio, while the features of system, battery audio are lower than the other three brands, while is the winner by the chief factors of price and cost. Customers have higher satisfaction to HW mobile phone at the feature of camera, screen, battery and signal than the other three brands. Meanwhile there are no obvious low level of satisfaction at other features. The Vi mobile phone has not particularly prominent and disadvantage, which all aspects are relatively balanced in the whole. In the conclusion, HW mobile phone

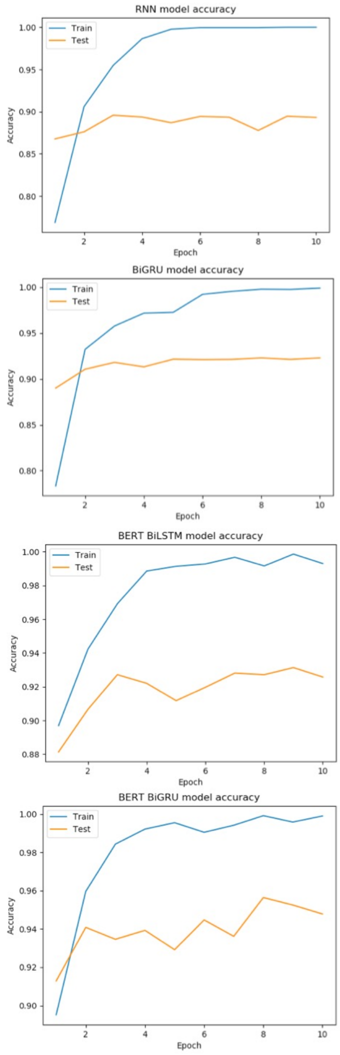

Figure 9: The Accuracy chart of RNN, BiGRU, Bert-BiLSTM and Bert-BiGRU Models

Table 8: Sentiment analysis results of phone products reviews by Bert-BiGRU-Softmax model

Brand	Attitude	Camera	Screen	System	Audio	Battery	Signal	Service	Price	Surface
AP	positive	0.734	0.549	0.784	0.551	0.679	0.389	0.726	0.521	0.695
	neutral	0.153	0.275	0.109	0.163	0.194	0.253	0.149	0.279	0.139
	negative	0.113	0.176	0.107	0.286	0.127	0.358	0.125	0.200	0.166
HW	positive	0.889	0.694	0.690	0.594	0.686	0.717	0.618	0.629	0.689
	neutral	0.094	0.172	0.189	0.159	0.159	0.156	0.106	0.194	0.171
	negative	0.017	0.134	0.121	0.247	0.155	0.127	0.276	0.177	0.140
Mi	positive	0.701	0.642	0.538	0.623	0.572	0.601	0.546	0.780	0.740
	neutral	0.106	0.183	0.224	0.151	0.179	0.213	0.193	0.117	0.169
	negative	0.193	0.175	0.238	0.226	0.249	0.186	0.261	0.103	0.091
Vi	positive	0.730	0.656	0.596	0.567	0.663	0.549	0.552	0.607	0.737
	neutral	0.142	0.185	0.207	0.161	0.163	0.332	0.177	0.208	0.137
	negative	0.128	0.159	0.197	0.272	0.174	0.119	0.271	0.185	0.126

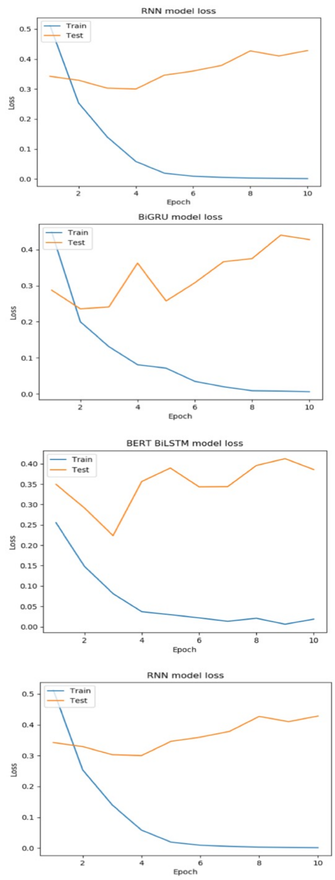

Figure 10: The Loss value map of RNN, BiGRU, Bert-BiLSTM and Bert-BiGRU models

has the highest degree of consumer satisfaction, AP brand is at second place of consumer satisfaction, Mi brand and Vi brand are at 3rd and 4th place, which the analysis result basically reflects the performance of four different mobile phones in accordance with the actual market conditions.

6. Conclusions

E-commerce reviews reveal the customers' opinions on the products and are helpful for manufacturer to improve their products quality selling on the e-commerce platforms. This paper proposed the deep learning algorithm of Bert-BiGRU-Softmax models to deal with the sentiment word disambiguation problem as well as the sentiment polarity annotation issue, which use the Bert model to extract features selection from E-commerce product reviews, the hidden layer of BiGRU model with attention mechanism to obtain semantic codes including the attention probability of input layer, and Softmax model to classify the sentiment tendency on the e-commerce product quality reviews. The experiment data set was based on the large-scale reviews from the Sunning, Taobao and other e-commerce websites. The experiment results show that the Bert-BiGRU-Softmax models have better performance than the RNN, BiGRU, Bert-BiLSTM, which improve the accuracy at least 3% on sentiment analysis of the e-commerce product quality reviews.

Declaration of competing interest

No author associated with this paper has disclosed any potential or pertinent conflicts which may be perceived to have impending conflict with this work.

Acknowledgments

This paper is supported by the Zhejiang Provincial Natural Science Foundation of China (Grant No. LY18G020009), the National Natural Science Foundation of China (Grant No. 71831006), the Zhejiang Provincial Philosophy and Social Science Planning Project of China (Grant No. 20NDJC096YB), the Hangzhou Science and Technology Project of China (Grant No. 20130142).

References

[1] Pak A, Paroubek P. " Twitter as a corpus for sentiment analysis and opinion mining," Proceedings of the 2010 International Conference on Language Resources and Evaluation. Paris: European Language Resources Association, pp.17-23,2010.

[2] Pavlou P A, Dimoka A, "The nature and role of feedback text comments in online marketplaces: Implications for trust building, price premiums, and seller differentiation," Information Systems Research, vol. 17, pp. 392-414 ,2006.

[3] Abbasi A, Chen H, Salem A, "Sentiment analysis in multiple languages: Feature selection for opinion classification in Web forums," ACM Transactions on Information Systems, vol. 26, pp. 12-21,2008.

[4] Zhang Z, Qiang YE, Yi-jun LI ," Literature review on sentiment analysis of online product reviews," J. Manag. Sci. China. vol.13, pp. 84-96,2010.

[5] Chang, C.C. and Lin, C.J. ," LIBSVM: A Library for Support Vector Machines," ACM Transactions on Intelligent Systems and Technology (TIST) , vol 2, pp. 1-27,2011.

[6] Hu F, Li L, Zhang Z L, et al. ," Emphasizing Essential Words for Sentiment Classification Based on Recurrent Neural Networks," Journal of Computer Science and Technology, vol.32, no. 4, pp.785-795, 2017.

[7] Aggarwal Charu C, Zhai Cheng Xiang." Mining Text Data," Springer New York Dordrecht Heidelberg London: Springer Science Business Media, LLC'12, 2012.

[8] Yelena Mejova, Padmini Srinivasan. "Exploring feature definition and selection for sentiment classifiers," In: Proceedings of the fifth international conference on weblogs and social media, Barcelona, Catalonia, Spain. pp. 17-21, 2011.

[9] Whitelaw Casey, Garg Navendu, Argamon Shlomo."Using appraisal groups for sentiment analysis," In: Proceedings of the ACM SIGIR Conference on Information and Knowledge Management (CIKM) , pp. 625-31 , 2005.

[10] Yu Liang-Chih, Wu Jheng-Long, Chang Pei-Chann, Chu Hsuan-Shou. ," Using a contextual entropy model to expand emotion words and their intensity for the sentiment classification of stock market news," Knowl-Based Syst, vol.41,pp. 89-97,2013.

[11] Jolliffe IT. ,"Principal Component Analysis ," Series: Springer Series in Statistics, 2nd ed. New York: Springer. ,2002.

[12] Duric Adnan, Song Fei. ,"Feature selection for sentiment analysis based on content and syntax models,". Decis Support Syst, vol.53 ,pp.704-11,2012.

[13] Griffiths Thomas L, Steyvers Mark, Blei David M, Tenenbaum Joshua B. " Integrating topics and syntax," Adv Neural Inform Process Syst ,pp.537-44,2005.

[14] Diana Maynard, Adam Funk. "Automatic detection of political opinions in tweets". In: Proceedings of the 8th international conference on the semantic web, ESWC'11, pp. 88-99,2011.

[15] Kim S, Hovy E.," Determining the Sentiment of Opinions," In Proc. the Conference on Computational Linguistics. Geneva, Switzerland, pp. 1367 -1373, 2004

[16] Turney P D, Littman M L. ," Measuring praise and criticism: Inference of semantic orientation from association,". ACM Transactions on Information Systems (TOIS), vol. 21, pp. 315-346, 2003.

[17] Y L Zhu, J Min, Y Q zhou. ," Semantic Orientation Computing Based on HowNet ,"Journal of Chinese Information Processing, vol. 20, pp. 14-20,2006.

[18] G. Somprasertsri, P. Lalitrojwong, " Mining feature-opinion in online customer reviews for opinion summarization ," Journal of Universal Computer Science, vol. 16,pp. 938-955, 2010.

[19] B LIU," Sentiment analysis and opinion mining(Synthesis Lectures on Human Language Technologies) ,". Morgan & Claypool Publishers, Vermont, Australia, vol. 5, pp. 1-167 ,2012.

[20] Asha S Manek, P Deepa Shenoy, et al. ," Aspect term extraction for sentiment analysis in large movie reviews using Gini index feature selection method and SVM classifier," World Wide Web, vol. 20, no.2 , pp.135-154, 2016.

[21] Pang B., Lee L.,Vaithyanathan S. ,"Thumbs up? sentiment classification using machine learning techniques ," In Proc. the Conference on Empirical Methods in Natural Language Processing, Philadelphia, USA, pp. 79-86, 2002.

[22] Go A, Bhayani R, Huang L. ," Twitter sentiment classification using distant supervision," CS224N Project Report, Stanford, pp. 1-12,2009.

[23] M Guerini, L Gatti, M Turchi,"Sentiment analysis: How to derive prior polarities from SentiWord Net," Breast Cancer Immunodiagnosis & Immunotherapy, pp. 3-11,2013.

[24] A Tripathy, A Agrawal, S K Rath ,"Classification of sentiment reviews using n-gram machine learning approach," Expert?Systems?with?Applications , vol. 57, pp. 117-126, 2016.

[25] KIM Y. ,"Convolutional neural networks for sentence classification," In Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing(EMNLP). Stroudsburg: Association for Computational Linguistics, pp. 1746-1751, 2014.

[26] Kyunghyun Cho, Bart van Merrienboer, Dzmitry Bahdanau ,"On the properties of neural machine translation: encoder-decoder approaches," In Proceedings of SSST-8, Eighth Workshop on Syntax, Semantics and Structure in Statistical Translation, Doha, Qatar, pp.103-111, 2014.

[27] Zhaowei Qu, Yuan Wang, Xaioru Wang," A Hierarchical attention network sentiment classification algorithm based on transfer learning,". Journal of Computer Applications, vol. 38, no.11, pp. 3053-3056, 2018.

[28] Trofimovich J. ,"Comparison of neural network architectures for sentiment analysis of Russian tweets," In Proceedings of the International Conference Dialogue 2016, RGGU ,2016.

[29] Pennington, J., Socher, R., and Manning, C. D.,"GloVe: Global vectors for word representation," Conference on empirical methods in natural language processing. ,2014.

[30] Devlin, J., Chang, M.W., Lee, K. and Toutanova, K. ,"BERT: Pre-Training of Deep Bidirectional Transformers for Language Understanding,". In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, pp.4171-4186, 2019.

[31] F. Abid, M. Alam, M. Yasir, C. Li,"Sentiment analysis through recurrent variants latterly on convolutional neural network of twitter," Future Generation Computer Systems, vol.95, pp. 292-308, 2019.

[32] Q. Qian, M. Huang, J. Lei, X. Zhu, "Linguistically regularized LSTMs for sentiment classification", in: Proc. 55th Annu. Meet. Assoc. for Comput. Linguist. vol. 1, pp. 1679-1689, 2016.

[33] L. Nio, K. Murakami," Japanese sentiment classification using bidirectional long short-term memory recurrent neural network", In Proceedings of the 24th Annual Meeting Association for Natural Language Processing, Okayama, Japan, pp. 1119-1122, 2018.

[34] Zhang, Min-Ling, and Zhi-Hua Zhou. "A review on multi-label learning algorithms, " IEEE transactions on knowledge and data engineering, vol. 26, no.8, pp.1819-1837, 2014.

[35] A. Hassan, A. Mahmood, "Convolutional recurrent deep learning model for sentence classification, " IEEE Access ,vol.6 , pp.13949-13957,2018.

[36] L. Mou, S. Member, P. Ghamisi, X.X. Zhu, S. Member, Corrections to "Deep recurrent neural networks for hyperspectral image classification", IEEE Trans. Geosci. Remote Sens., vol. 56, no.2 , pp. 1214-1215, 2018.

[37] Bahdanau, Dzmitry, Cho, Kyunghyun, Bengio, Yoshua, "Neural Machine Translation by Jointly Learning to Align and Translate," In Proceedings of the International Conference on Learning Representations 2015 (ICLR2015),San Diego, CA, USA, pp.1-15,2014.

[38] Dong Li, Wei Furu, Tan Chuanqi, et al. ," Adaptive recursive neural network for target-dependent Twitter sentiment classification," Baltimore: Assoc. Computational Linguistics-ACL, pp. 49-54,2014.

[39] S.M. Gru?.er-Sinopoli, C.N. Thalemann," Bidirectional recurrent neural networks," IEEE Trans. Signal Process, vol. 45, no.11, pp.2673-2681, 1997.

[40] Liu J, Zhang Y. ,"Attention modeling for targeted sentiment ,"In Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: vol. 2, pp572-577,2017.

[41] C.P. Andrew, L. Maas, Raymond E. Daly, Peter T. Pham, Dan Huang, Andrew Y. Ng, ”Learning word vectors for sentiment analysis,” in: Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies, pp. 142-150, 2011.

[42] K. Saranya, S. Jayanthy, ”Onto-based sentiment classification using machine learning techniques,” in: Proceedings of 2017 International Conference on Innovations in Information, Embedded and Communication Systems, ICIIECS 2017, pp. 1-5, 2017.

[43] Songbo Tan, Jin Zhang. ”An empirical study of sentiment analysis for Chinese documents,” Expert Systems With Applications, vol. 34, no.4, pp.2622-2629, 2008.

[44] Weijiang Li, Fang Qi. ”Sentiment analysis based on multi-channel bidirectional long short term memory network,” Journal of Chinese Information Processing, vol.33, no.12, pp.119-128,2019.

[45] Yujuan Yang, Huanhuan Yuan, Yongli Wang. ”Sentiment analysis method for comment text,” Journal of Nanjing University of Science and Technology, vol.43, no.3, pp.280-285, 2019.