



*Research article*

## **Sentiment analysis for e-commerce product reviews by deep learning model of Bert-BiGRU-Softmax**

**Yi Liu\*, Jiahuan Lu, Jie Yang and Feng Mao\***

Management School, Hangzhou Dianzi University, Hangzhou, 310018, China

\* **Correspondence:** Email: liuyi@hdu.edu.cn; fengmao@hdu.edu.cn.

**Abstract:** Sentiment analysis of e-commerce reviews is the hot topic in the e-commerce product quality management, from which manufacturers are able to learn the public sentiment about products being sold on e-commerce websites. Meanwhile, customers can know other people's attitudes about the same products. This paper proposes the deep learning model of Bert-BiGRU-Softmax with hybrid masking, review extraction and attention mechanism, which applies sentiment Bert model as the input layer to extract multi-dimensional product feature from e-commerce reviews, Bidirectional GRU model as the hidden layer to obtain semantic codes and calculate sentiment weights of reviews, and Softmax with attention mechanism as the output layer to classify the positive or negative nuance. A series of experiments are conducted on the large-scale dataset involving over 500 thousand product reviews. The results show that the proposed model outperforms the other deep learning models, including RNN, BiGRU, and Bert-BiLSTM, which can reach over 95.5% of accuracy and retain a lower loss for the e-commerce reviews.

**Keywords:** sentiment analysis; e-commerce product reviews; Bert; Bidirectional GRU

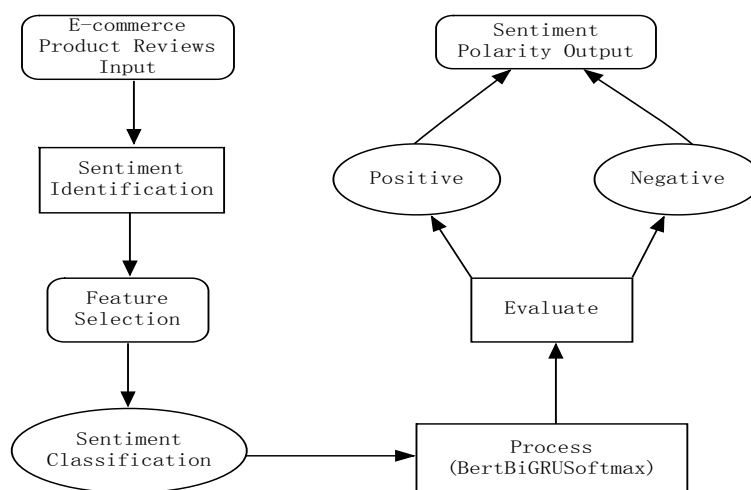
---

### **1. Introduction**

With the rapid growth of e-commerce, online shopping can bring convenience and lower price products for consumers. Due to the information inconsistency between the real product quality and description provided by the seller, more and more consumers seek product information from e-commerce reviews involving many features (dimensions) of price, service, logistics and so on. E-commerce reviews plays a key role for customers in shopping decision process, meanwhile the manufacturers and E-commerce platforms can improve their products' quality and on-line services

based on the public opinions [1]. 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 content analysis technique to distinguish the positive or negative attitudes of consumer from the related information appearing in different forms like BBS, blogs, Wiki or forum websites (Abbasi et al., 2008) [3].

The challenges of sentiment analysis (SA) of E-commerce product reviews lie in dimensions mapping and sentiment word disambiguation. The dimension mapping problem refers to mapping opinioned text blocks with right dimensions. The sentiment word disambiguation problem refers the situation that a sentiment word may be connected with two or more dimensions. Therefore, sentiment analysis of E-commerce reviews is considered as a multi-dimensional classification process. As shown in Figure 1, it has three main classification levels: document-level, sentence-level, and aspect-level tasks. Document-level SA aims to classify opinion document as expressing 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 relative 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. The second approach is machine learning, which regards sentiment analysis as a binary or multi-class classification task and uses common classification methods (Chang and Lin, 2011) [5]. The latest sentiment analysis method is the deep learning models including the Convolutional Neural Network (CNN), Recurrent Neural Network (RNN) and other models (Hu and Li, 2017) [6].



**Figure 1.** Sentiment analysis process of E-commerce product reviews.

## 2. Related work

### 2.1. Feature selection

The first step of sentiment analysis is to extract text features. The main feature selection methods are as following.

Terms presence and frequency: It either gives the words binary weighting (zero when the word appears, or one at other condition) or uses term frequency weights to indicate the relative importance of features, which is based on the frequency counts of individual words or word n-grams [7].

Parts of speech (POS): It is to find out important indicators of opinions such as adjectives, adverb and so on.

Opinion words and phrases: These words are usually used to express opinions including positive or negative. On the other hand, some phrases express opinions without any emotional words.

Feature selection attempts to reduce the dimension of data by picking from the original attribute set, 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 and bootstrap this set through synonym detection or on-line resources to obtain larger lexicon [8]. The mutual information measure provides a formal way to model the mutual information between the features and the classes. Turney and Littman (2003) use 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 [9] outperforms the PMI-based expansion methods by developing contextual entropy model to expand the set of seed words generated from a small corpus of stock market news. Their method combines both co-occurrence strength and contextual distribution, and need more useful emotion words and fewer noisy words.

Feature transformation methods create the smaller set of features as the function of the original feature set. Latent Semantic Indexing (LSI) method transforms the text space into a new axis system which is the linear combination of original word features. Principal Component Analysis techniques (PCA) are used to achieve optimal goal [10]. There are other statistical approaches which could be used in Feature Selection (FS) like Hidden Markov Model (HMM) and Latent Dirichlet Allocation (LDA). Duric and Song [11] apply it 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 the collection of documents [12].

## 2.2. Sentiment classification

The second step of sentiment analysis is sentiment classification. Sentiment classification can be divided into machine learning, lexicon based approach and deep learning approaches [13].

Kim and Hovy (2004) analyzed the emotional tendency of text based on the synonyms, antonyms and hierarchies of the WordNet dictionary [14]. The acquisition of sentiment tendency about Chinese text mainly depended on HowNet [15]. YL 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 [16]. G. Somprasertsri (2006) extracted the product feature and opinion of online customer reviews based on the consideration of syntactic and semantic information [17]. Sentiment analysis model of polarity lexicons is commonly used, but the existing sentiment vocabulary lexicons is limited (B Liu, 2014) [18].

Compared with sentiment polarity lexicon, Machine learning methods have better advantages in the nonlinear and high dimensional pattern recognition problem. Pang (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% [19]. Go and Bhayani (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 [20]. Marco Guerini (2013) applied 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% [21]. 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, experimental results show that the SVM combined unigram, bigram and trigram features achieved the best classification results [22].

With the development of deep learning research, the deep neural network has outstanding performance in natural language processing. Kim (2014) used Convolutional Neural Network (CNN) to solve sentiment classification problem [23]. Cho (2014) proposed Gated Recurrent Unit (GRU) to analyze long dependent contexts, which has significant improvements in various tasks [24]. Qu and Wang (2018) proposed the sentiment analysis model based on hierarchical attention network, which improved by 5% accuracy compared with the recurrent neural network [25]. Due to the reviews become larger, the traditional neural network models can't completely capture the entire context of sentence or review, which the novel neural networks are needed by changing the different types of word embedding (i.e., GloVe or BERT) [26]. There are variants of RNNs exploited for sentiment analysis, such as Bidirectional-LSTM (BiLSTM) and Bidirectional-GRU (BiGRU) [27]. Trofimovich (2016) used LSTM (long short-term memory) to solve sentiment analysis problem [28]. The LSTM model was offered to make phrase-level sentiment classification centered on the regularization containing linguistics such as negativity, intensity and polarity [29]. The bidirectional LSTM is used to train the tagging text in order to handle syntactic and semantics of the Japanese language [30]. The BiGRU model with multi-input and output was proposed for phrase-level polarity classification [31].

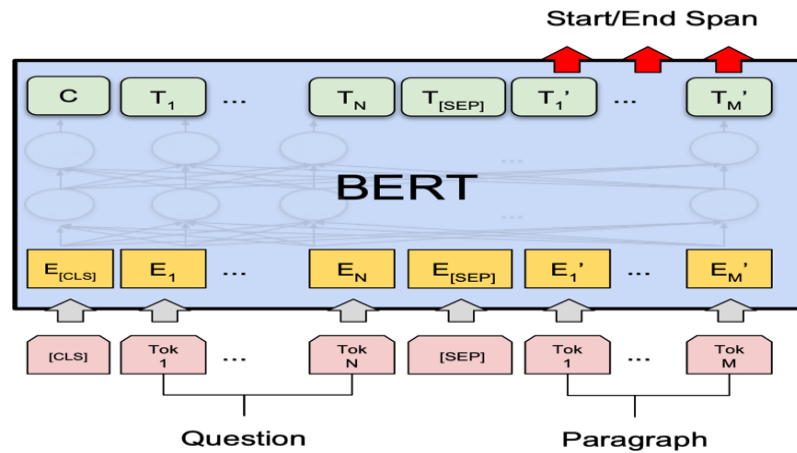
This paper researches the deep learning model of Bert-BiGRU-Softmax for sentiment analysis of the E-commerce product reviews, in which the Bert model makes the features exaction by mapping each sentence into the proper dimension, Bi-GRU model solves the sentiment word disambiguation problems by using the multiple clause-rules recognition, and the Softmax model with attention mechanism makes the multi-dimensional sentiment analysis by calculating the sentiment polarity at 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 (Devlin, 2019) [32]. The BERT applies the bidirectional training of transformer to learn contextual relations between words and tokens ( $Tok_1, \dots, Tok_M$ ) in text. The transformer includes two separate mechanisms that an encoder ( $E_1, \dots, E_N$ ) reads the text input and a decoder that makes 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 ( $T_1, \dots, T_N$ ) 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. Figure 2 shows the multi-level structure chart of BERT model.

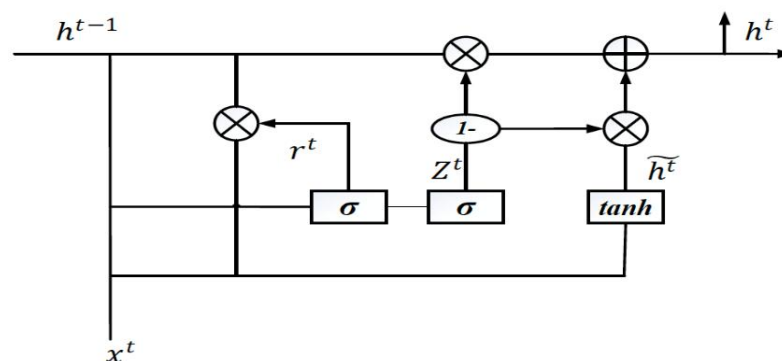


**Figure 2.** The multi-level structure of Bert model.

BERT alleviates the bidirectional constraints by using “masked language model” (MLM) pre-training natural language process. The masked language model randomly masks some of the tokens from the input, and the objective is to predict the original vocabulary of the masked word based on its context. This bidirectional training model allows the BERT have deeper sense of language context and learn the context of word by all of its surrounding.

### 3.2. GRU model

The GRU is a specific model of recurrent neural network that uses 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 following 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 [33]. The structure of GRU is shown as Figure 3.



**Figure 3.** The structure of GRU model.

As the refinement of recurrent neural network, GRU has what's called an update gate  $z_t$  and a reset gate  $r_t$ . Using input vectors  $x_t$  and output vectors  $h_t$ , the model refines outputs by controlling the  $h_{t-1}$  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 way to describe these technologies as the “memory-centered” type of neural network. By contrast, other types of neural networks without gated recurrent units often don't have the ability to retain information.

### 3.3. Attention mechanism

Without attention mechanism, translation relies on reading a complete sentence and compressing 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. Bahdanau (2014) proposes attention mechanism to memorize the long source sentences in Neural Machine Translation (NMT) and NLP [34]. 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 information that 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 model is shown as Figure 4.

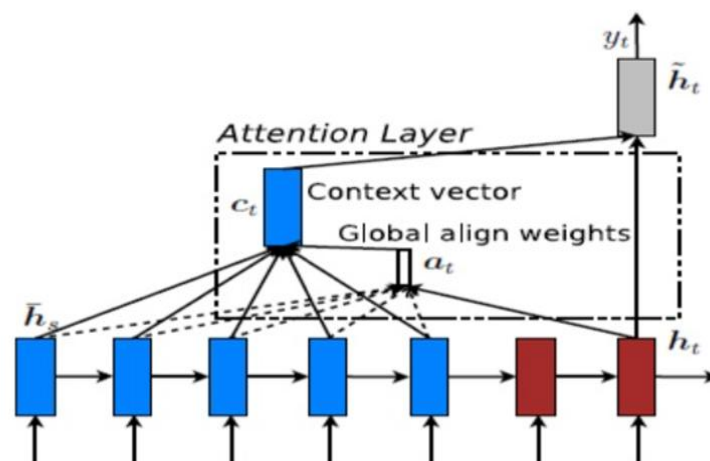
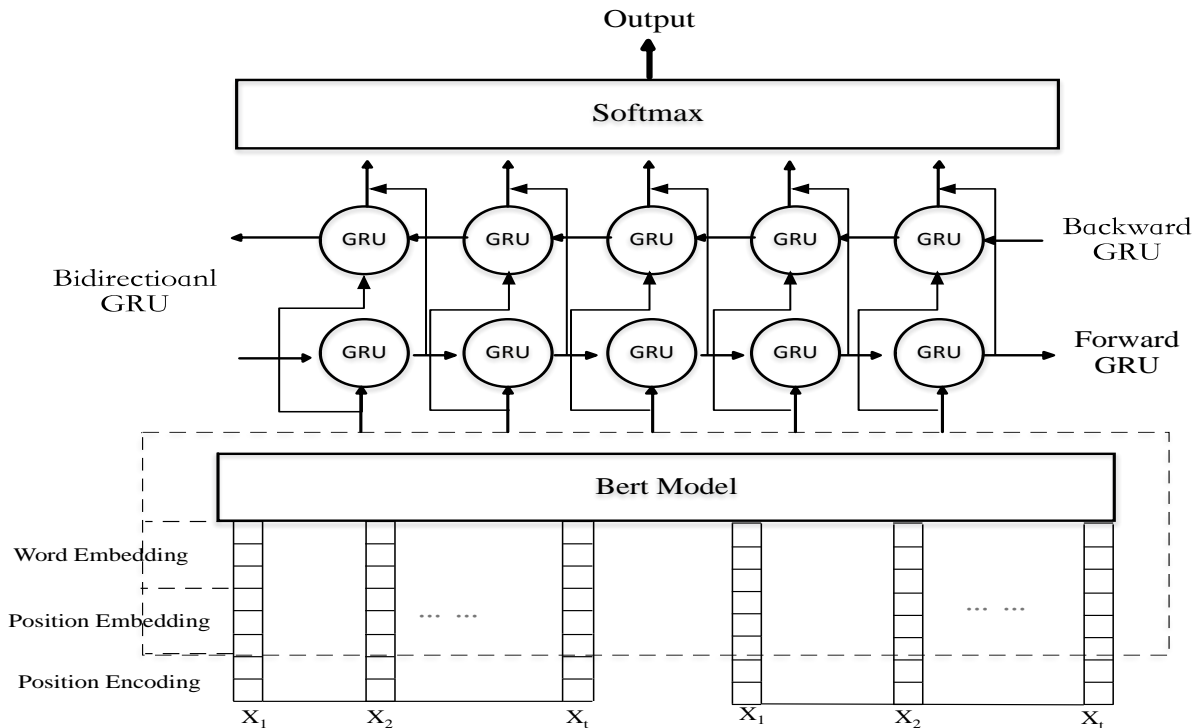


Figure 4. The structure of Attention Mechanism model.

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

This paper researches Bert-BiGRU-Softmax model for sentiment analysis on the e-Commerce product quality reviews. The sentiment Bert model, as input layer, makes the feature extraction at pre-processing-phase. The hidden layer of Bidirectional GRU performs 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. The output layer of Softmax calculates the sentiment polarity by pooling to smaller weighted dimensions according to the attention mechanism. The Figure 5 is the structure chart of the Bert-BiGRU-Softmax model.



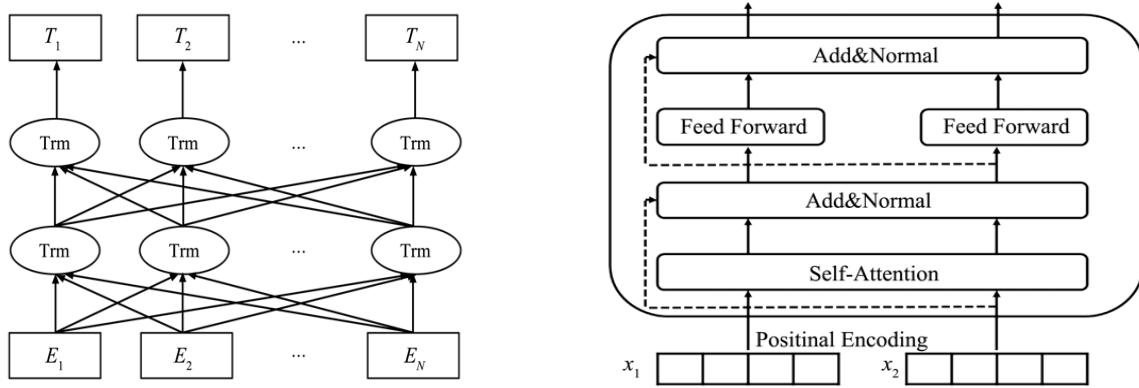
**Figure 5.** Information flow chart of the Bert-BiGRU-Softmax model.

#### 4.1. Sentiment-Bert

Dimensions mapping of the e-commerce product quality reviews are typically domain dependent problem. Since the 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.

The Bert transforms the e-commerce reviews into a numerical matrix, where each column represents the identified features and each row expresses a particular review. The matrix is given as input of Bert in order to train the model by two training strategies of the MLM and next sentence prediction, with the best performance of the combined loss function [35]. When training the BERT model, sentences segmentation procedure is used for splitting a long sentence into several short text blocks, and word sequences are regarded as nodes of text blocks, which is more efficiently to map those short sentences into corresponding dimensions. The transformer receives the pairs of sentence 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 as Figure 6.

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



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

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  $D_c$  of category  $c$  directly. For each word  $w_i$  ( $w_1, w_2, w_3, \dots, w_m$ ) in the review sentence  $S$  and a dimension  $d_j$ , 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 Eq (1)).

$$p(S) = p(w_1, w_2, \dots, w_m) = \prod_{i=1}^m p(w_i | w_1, w_2, \dots, w_{i-1}) \quad (1)$$

Transformer add sequence information by word positional embedding (PE) Eqs (2) and (3).

$$PE_{(\text{pos}, 2i)} = \sin(\text{pos}/10000^{2i/d_{\text{model}}}) \quad (2)$$

$$PE_{(\text{pos}, 2i+1)} = \cos(\text{pos}/10000^{2i/d_{\text{model}}}) \quad (3)$$

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  $x_{[CLS]}$  and  $x_{[MASK]}$  from the special words of  $w_{[CLS]}$  and  $w_{[MASK]}$  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 are represented as Eqs (4,5).

$$\text{loss} = \sum_i L(x_{[MASK]}^i) F(x_{[MASK]}^i) \quad (4)$$

$$F(x_{[MASK]}^i) = \begin{cases} k & \text{if } x \in R \\ 1 & \text{if } x \notin R \end{cases} \quad (5)$$

where  $R$  is the emotional words lexicon of e-Commerce product quality,  $L(x_{[MASK]}^i)$  is the loss of the masked words  $x_{[MASK]}^i$  in prediction process of input sequence sentence,  $F(x_{[MASK]}^i)$  is the weight of emotional and non-emotional words. Finally we map a review sentence  $S$  with the dimension  $d_j$  that has the maximum probability score with  $S$ . The sentiment Bert model for Sentence information processing 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 by 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.

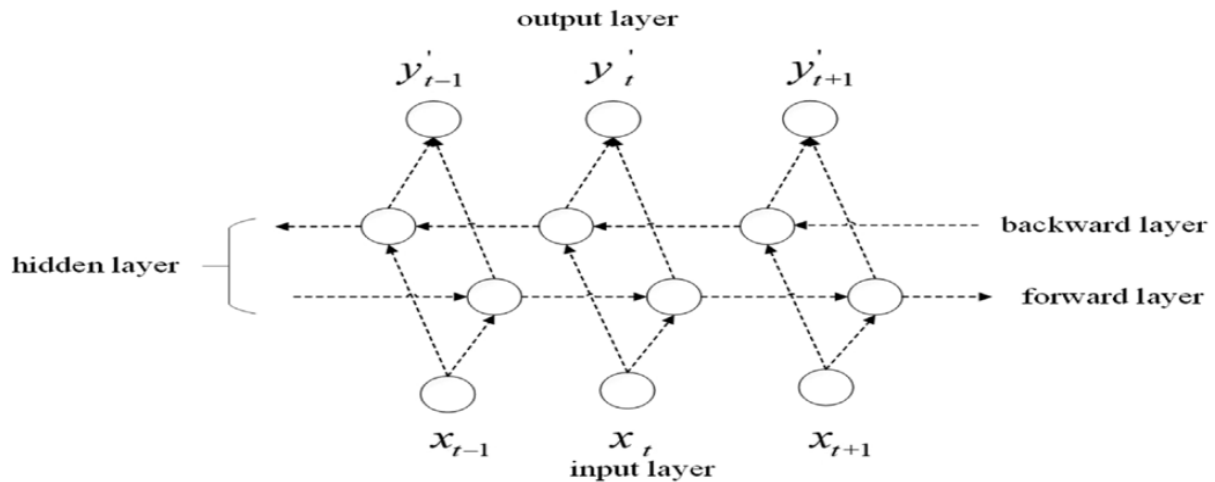
End

---

#### 4.2. Bidirectional GRU With Clause-Rules recognition

Sentiment lexicons of the sentiment polarity in e-commerce reviews are widely used in sentiment analysis. However, those lexicons did not distinguish the polarities by one word on different dimensions of product. This paper uses multiple rules to enhance the performance of sentiment analysis. For example, when sentences with positive strong rule, “It is so famous brand of mobile phone that the customer would like to buy Huawei.”, which is labelled as positive without other factors. The emotion analysis has the relationship with the previous state and next state of the text sequence at the same time. GRU is one-way neural network structure of this fixed-length context vector design, which will forget the part information during the whole input context process.

This paper proposes 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 [36]). The idea of BiGRU is to split the regular GRU neurons into a forward state and a backward state. 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. Therefore, the BiGRU is able to learn the information from two different directions, and make more accurate analysis of the e-commerce product quality reviews based on clause-rules recognition rules. The structure of BiGRU is shown in Figure 7.



**Figure 7.** The structure of Bidirectional GRU network.

BiGRU has the given sequence of input vectors  $\langle x_1, x_2, \dots, x_t \rangle$  ( $x_t$  represents the concatenations of input features), setting unit computes the corresponding hidden activations  $\langle h_1, h_2, \dots, h_t \rangle$ , and outputs a vector sequence  $\langle y_1, y_2, \dots, y_t \rangle$  from the input data. At time  $t$ , the current hidden state of  $h_t$  is determined by three parts including input vectors  $\langle x_1, x_2, \dots, x_t \rangle$ , the forward hidden state  $h_t$  and the backward hidden state  $h_t$  at the same 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.  $\sigma$  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 by Eqs (6)–(10).

$$z_t = (w_z[h_{t-1}, x_t]) \quad (6)$$

$$r_t = (w_r[h_{t-1}, x_t]) \quad (7)$$

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

$$h_t = GRU(x_t, h_{t-1}) \quad (9)$$

$$h_t = w_t h_t + v_t h_t + b_t \quad (10)$$

$w_z$  and  $w_r$  are the weight matrices vectors 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  $t$ . the symbol  $*$  denotes the element-wise multiplication, and the symbol  $[]$  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  $t$  by 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  $t$  by the following steps:

---

- (1) The input  $x_t$  is multiplied by a weight  $w_{xt}$ ;
- (2) The forward output  $h_{t-1}$  which hold forward units information multiplied by weight  $w_{hr}$ ;
- (3) Both are added together and sigmoid function  $[0, 1]$  apply to squeeze the output.

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  $h_{t-1}$ ; this allows to pass only the relevant past information.
- (3) Both are added together and  $\tanh$  function is applied.

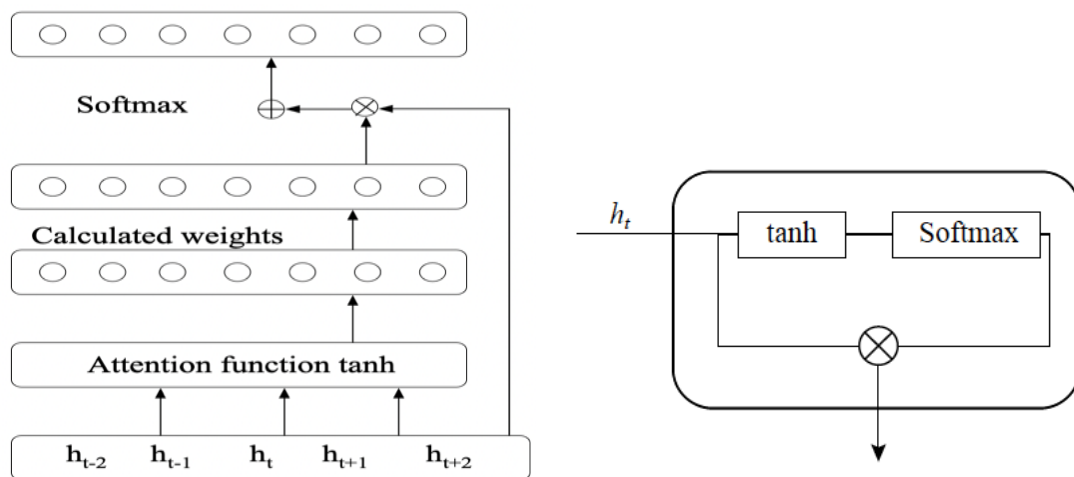
Final memory at current time step: Last but not least the unit has to calculate the  $h_t$  vector which holds information for the current unit and it will pass it further down to the network. Key role in this process plays the update gate  $z_t$ .

- (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

### 4.3. Softmax with attention mechanism

The output layer of the Softmax makes 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, which the attention probability and sentiment polarity weights of reviews are calculated by the Softmax layer(as shown in Figure 8).



**Figure 8.** The flow chat of Softmax with attention mechanism.

The detailed approaches of the Softmax with attention mechanism [37] are as following:

- (1) After receiving the feature vector of attention layer, attention function calculates the weight matrix  $w_a$ :

$$u_t = \tanh(w_a h_t + b_a) \quad (11)$$

$$\alpha_t = \frac{\exp(u_t^T u_a)}{\sum_t \exp(u_t^T u_a)} \quad (12)$$

$$s_{it} = \sum_{i=1}^n a_{it} h_{it} \quad (13)$$

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  $\alpha_t$  is computed as average weight.

(2) Make the normalization operation of Softmax by Eq (14):

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

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

## 5. Experimental design and result analysis

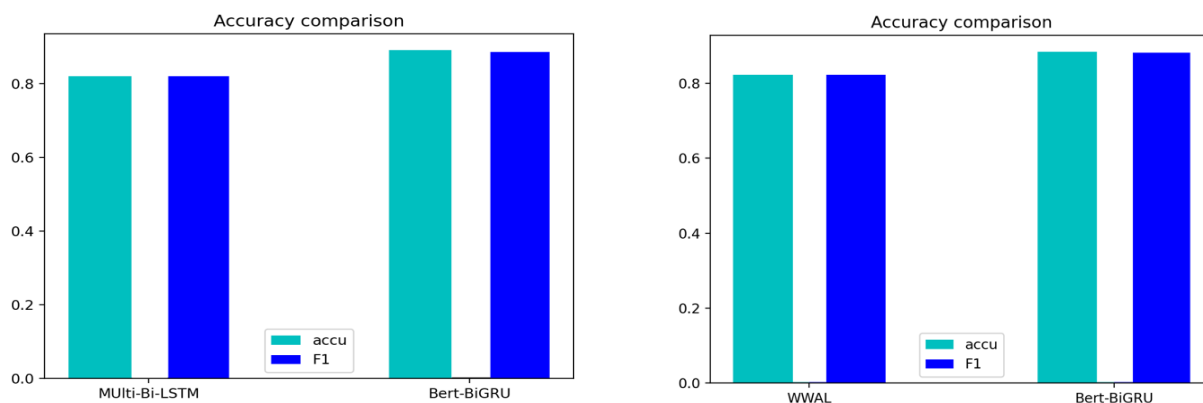
### 5.1. Experiment 1

In order to evaluate the sentimental analysis efficiency of the Bert-BiGRU-Softmax Model, this paper conduct experiments on multi-source datasets with various domains by the Python 3.5 and Tensorflow framework at PyCharm IDE. The first dataset is IMDB, which is used and prolonged extensively as a benchmark dataset [38]. The dataset contains 25,000 tweets with the polarity of 12,500 positives, and 12,500 negatives about the movie reviews. The second dataset is the Chinese emotional corpus known as ChnSentiCorp [39], which has abundant Sentiment Corpus including ChnSentiCorpHtl, ChnSentiCorpMov and so on. Table 1 shows the experimental results of proposed emotional analysis model, the multi-Bi-LSTM model and the Weight W2V-ATT -LSTM (WWAL) model on ChnSentiCorp-Htl-ba-6000 and COAE2014-task4.

**Table 1.** The Accuracy and F-score of different models.

Datasets	Model	Accuracy	F1
COAE2014-task4	CNN[40]	0.8820	0.8760
	Multi-Bi-LSTM[40]	0.8840	0.8840
	Bert-BiGRU	0.8903	0.8857
ChnSentiCorp-Htl-ba-6000	WWAL(WeightW2V-ATT-LSTM) [41]	0.8727	0.8719
	Bert-BiGRU	0.8831	0.8816

In first experiment of COAE2014-task4, the accuracy and F1 measure of Bert-BiGRU-Softmax is (0.8903, 0.8857), which is 1% larger than Multi-Bi-LSTM (0.8840, 0.8840). The second experiment of ChnSentiCorp-Htl-ba-6000, Bert-BiGRU-Softmax performs better accuracy (0.8831) and F1 measure (0.8816), which are slightly higher than accuracy (0.8727) and F1 measure (0.8719) of WWAL (as shown in Figure 9). When the feature classifier transfers from different topics, the machine learning techniques of sentiment classifier is severely dependent on its domains. The Multi-Bi-LSTM, Weight W2V-ATT-LSTM and Bert-BiGRU-Softmax models are suitable for sentiment analysis, which Bert-BiGRU-Softmax has better performance by Bert model as pretraining reviews feature extractor.



**Figure 9.** The Accuracy of WWAL, Multi-Bi-LSTM and Bert-BiGRU-Softmax.

## 5.2. Experiment 2

This paper crawls and analyzes a large-scale dataset with 150 predefined dimensions of 500 thousands E-commerce reviews about mobile phone products from Sunning, Taobao and other e-commerce websites, which cover almost all the aspects of different products with their polarity concerning positive and negative, handling the dimensions “quality”, “logistic” and “service”. We use the deep learning models of RNN, LSTM, GRU, BiGRU, BiLSTM, Bert-BiLSTM and Bert-BiGRU-Softmax to make sentiment analysis of e-commerce product quality reviews from dimensions of brand, ratings, price and others. Table 2 is 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 new brand phones. Also, the highest score for “trust” among all the emotions shows that the reviewers are conviction and trust the product.

**Table 2.** The sentiment reviews about mobile phone quality on e-Commerce websites.

---

This mobile phone provided bigger screen and powerful battery, the only drawback for me is that it’s a little thick ...

---

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!

---

This mobile phone has wonderful software function and appearance which could be better. My friend recommended it to me, Although charge is...

---

I admit, mobile phone has good camera system with high pixel. <br/>If anything, it helps you make beautiful picture what you want<br/>.

---

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. ...

---

The particularly nice mobile phone is received, the audio sounds good! courier delivery it to my home, give five star rating for it ...

---

This paper compares the performances of RNN, LSTM, GRU, BiGRU, BiLSTM, Bert-BiLSTM and Bert-BiGRU-Softmax models according to the accuracy and loss value, which are based on the

400 thousands original data from the web reviews corpus, and split into training and test datasets in experiments. The hyper parameters of Bert-BiGRU-Softmax model are as following:

**Table 3.** The hyper parameters of Bert-BiGRU-Softmax model.

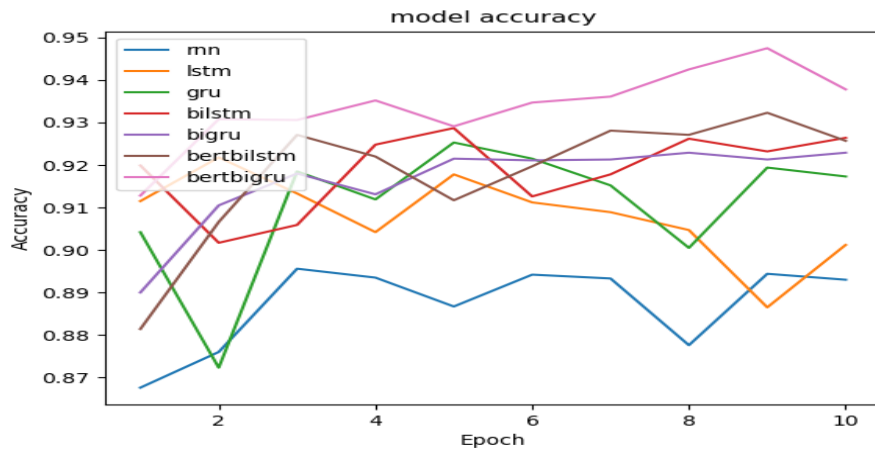
Hyper parameter	Value
The nodes of hidden layer network	100
Loss function	Categorical_Crossentropy
Optimizer	Adam
Dimensions of word vector	768
Epoch	10

The accuracy and loss value of models on the different datasets are shown in Tables 4 and 5. The accuracy and loss cure of models on the different datasets are shown as Figures 10 and 11.

According to the accuracy cure, the RNN model reaches the highest accuracy value of 0.895 at the 3rd epoch, and then the accuracy value of RNN becomes volatility at the over-fitting status. The BiGRU model reaches the highest accuracy value of 0.922 at the 8th epoch, and then the accuracy value ascends to the over-fitting status. The LSTM model reaches the highest accuracy value of 0.921 at the 2th epoch. The GRU model becomes over-fitting status after getting highest accuracy of 0.925 at the 5th epoch. The BiLSTM model reaches the highest accuracy value of 0.928 at the 5th epoch. The Bert-BiLSTM model reaches the highest accuracy value of 0.931 at the 9th epoch, and then the accuracy value of LSTM descend to the over-fitting status. Comparing with the BiGRU model, the Bert-BiGRU-Softmax model reaches the highest accuracy value of 0.955 at the 8th epoch and reduces computational overhead, which is superior to the competitor algorithms.

**Table 4.** The accuracy of different deep learning models.

Model	Training Epochs									
	1	2	3	4	5	6	7	8	9	10
RNN	0.867	0.876	0.895	0.893	0.886	0.894	0.893	0.877	0.894	0.893
LSTM	0.911	0.921	0.913	0.904	0.917	0.911	0.908	0.904	0.886	0.901
GRU	0.904	0.872	0.918	0.911	0.925	0.921	0.915	0.901	0.919	0.917
BiLSTM	0.919	0.901	0.905	0.924	0.928	0.912	0.917	0.926	0.923	0.926
BiGRU	0.890	0.910	0.918	0.913	0.921	0.921	0.921	0.922	0.921	0.922
Bert-BiLSTM	0.881	0.906	0.927	0.922	0.911	0.919	0.928	0.927	0.931	0.925
Bert-BiGRU	0.912	0.940	0.934	0.939	0.929	0.944	0.936	0.955	0.952	0.947



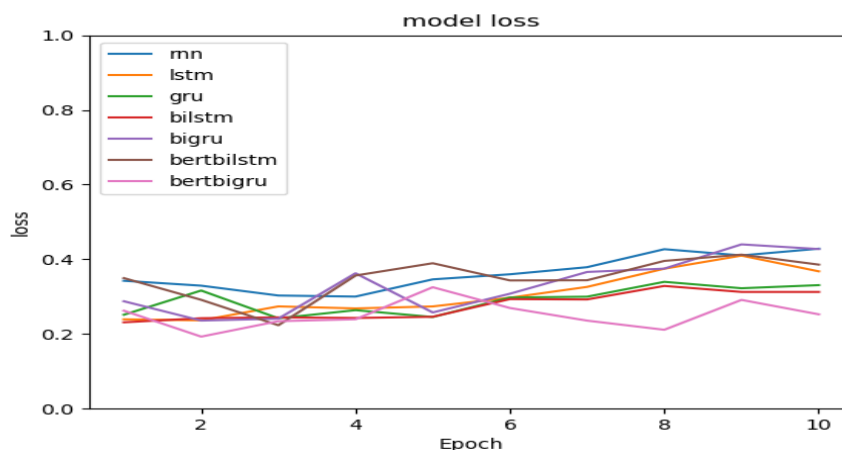
**Figure 10.** The Accuracy of RNN, LSTM, GRU, BiGRU, Bert-BiLSTM and Bert-BiGRU.

**Table 5.** The loss value of different deep learning model.

Model	Training Epochs									
	1	2	3	4	5	6	7	8	9	10
RNN	0.342	0.329	0.303	0.300	0.346	0.359	0.378	0.427	0.410	0.428
LSTM	0.239	0.236	0.273	0.268	0.273	0.297	0.326	0.375	0.409	0.367
GRU	0.251	0.316	0.242	0.263	0.245	0.297	0.299	0.339	0.322	0.331
BiLSTM	0.231	0.242	0.244	0.243	0.246	0.293	0.292	0.328	0.313	0.313
BiGRU	0.287	0.235	0.240	0.362	0.257	0.307	0.366	0.374	0.439	0.427
Bert-BiLSTM	0.349	0.291	0.223	0.356	0.389	0.343	0.343	0.395	0.412	0.385
Bert-BiGRU	0.262	0.192	0.234	0.239	0.325	0.269	0.235	0.211	0.291	0.252

According to the loss curve of different models, the extraction ability of RNN is the worst of all models. Comparing with LSTM model, GRU model has better performance. BiGRU model will generate large computational overhead and poor accuracy performance than BiLSTM model. Bert-BiLSTM model can reach higher performance than that of BiGRU. The proposed Bert-BiGRU-Softmax model reaches the highest accuracy value of 0.955, which is improved by nearly 3.6% of Bert-BiLSTM at the training datasets. By comparing the models from different types of comment sets, experiments show that the Bert-BiGRU-Softmax model can change dimension flexibly and improve the accuracy of feature extraction.

Bert-BiGRU-Softmax model is applied for the sentiment analysis of mobile phone product quality reviews, which obtains each type of polarities (positive, neutral and negative) including the attention probability distribution from input text sequences and calculates the weights of customer reviews emotional scores according to the different dimensions (camera, screen, system, audio, battery, signal, service, price and surface). The results of sentiment analysis are shown in Table 6.



**Figure 11.** The Loss of RNN, LSTM, GRU, BiGRU, Bert-BiLSTM and Bert-BiGRU.

**Table 6.** Sentiment analysis results of phone 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

After the comprehensive compared, AP mobile phone gets higher satisfaction at the features of system, hardware and service than the other three brands from consumers, while gets the lowest satisfaction of signal and price problem in the four brand products. Mi mobile phone has the highest satisfaction from consumers at the feature of price, surface and audio, while the features of system, battery audio are lower than the other three brands. Customers have higher satisfaction to HW mobile phone at the features of camera, screen, battery and signal than the other three brands. Meanwhile there is 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 conclusions, HW mobile phone has the highest satisfaction degree from consumer, which AP, Mi and Vi brand are at 2nd, 3rd and 4th place according to the consumer satisfaction. The analysis results reflect the performance of four different brand mobile phones in accordance with the actual market conditions.



## 6. Conclusion

E-commerce reviews reveal the customers' opinions about the products selling at the e-commerce platforms, which are helpful for manufacturers to improve their products quality. This paper proposes the deep learning model of Bert-BiGRU-Softmax to deal with the sentiment word disambiguation and the sentiment polarity problem, which uses the Bert model to extract features selection from e-commerce reviews at input layer, the hidden layer of BiGRU model with attention mechanism to obtain semantic codes including the attention probability of input layer, and Softmax model as the output layer to classify the sentiment tendency of the e-commerce reviews. The experiment analyses the large-scale review datasets from the Sunning, Taobao websites. The experimental results show that Bert-BiGRU-Softmax model has better performance than RNN, BiGRU and Bert-BiLSTM, which improves the accuracy at least 3% on sentiment analysis of the e-commerce product quality reviews.

## Acknowledgments

This work is supported by the Zhejiang Provincial Natural Science Foundation of China (Grant No. LY18G020009), the Zhejiang Provincial Philosophy and Social Science Planning Project of China (Grant No. 20NDJC096YB), the National Key Research and Development Project of China (Grant No. 2017YFF0209605). We are thankful to Feng Qian and Hu Gu providing expertise and assistance in the research, who work in quality inspection and technical supervision institution of Hangzhou in China.

## Conflict of interest

The authors declare no conflict of interest.

## References

1. P. Sasikala, L. M. I. Sheela, Sentiment analysis of online product reviews using DLMNN and future prediction of online product using IANFIS, *J. Big Data*, **7** (2020), 33–53.
2. P. A. Pavlou, A. Dimoka, The nature and role of feedback text comments in online marketplaces: Implications for trust building, price premiums, and seller differentiation, *Inf. Syst. Res.*, **17** (2006), 392–414.
3. A. Abbasi, H. Chen, A. Salem, Sentiment analysis in multiple languages: feature selection for opinion classification in web forums, *ACM Trans. Inf. Syst.*, **26** (2008), 12–21.
4. Z. Zhang, Y. E. Qiang, Literature review on sentiment analysis of online product reviews, *J. Manage. Sci. China*, **13** (2010), 84–96.
5. C. Chang, C. J. Lin, LIBSVM: A library for support vector machines, *ACM Trans. Intel. Syst. Technol.*, **2** (2011), 1–27.
6. F. Hu, L. Li, Z. L. Zhang, et al, Emphasizing essential words for sentiment classification based on recurrent neural networks, *J. Comput. Sci. Technol.*, **32** (2017), 785–795.
7. Y. Mejova, P. Srinivasan, Exploring feature definition and selection for sentiment classifiers. In *Proc. 5<sup>th</sup> Int. Conf. weblogs social media*, Barcelona, Catalonia, Spain, (2011), 17–21.
8. W. Casey, G. Navendu, A. Shlomo, Using appraisal groups for sentiment analysis. In *Proc. ACM SIGIR Conf. Inform. Knowl. Manag.*, (2005), 625–31.

9. L.-C. Yu, H.-L. Wu, P.-C. Chang, H.-S. Chu, Using a contextual entropy model to expand emotion words and their intensity for the sentiment classification of stock market news, *Knowl. Based Syst.*, **41** (2013), 89–97.
10. I. Jolliffe, N. A. Jolliffe, *Principal Component Analysis. Springer Series in Statistics*, 2nd edition, Springer-Verlag, New York, 2002.
11. D. Adnan, F. Song, Feature selection for sentiment analysis based on content and syntax models, *Decis. Support Syst.*, **53** (2012), 704–11.
12. L. G. Thomas, S. Mark, M. B. David, B. T. Joshua, Integrating topics and syntax, *Adv. Neural Inf. Process. Syst.*, (2005), 537–544.
13. A. Yadav, D. K. Vishwakarma, Sentiment analysis using deep learning architectures: a review, *Artif. Intell. Rev.*, **53** (2020), 4335–4385.
14. S. Kim, E. Hovy, Determining the sentiment of opinions. In *Proc. Conf. Comput. Linguis.*, Geneva, Switzerland, (2004), 1367–1373.
15. P. D. Turney, M. L. Littman, Measuring praise and criticism: Inference of semantic orientation from association, *ACM Trans. Inf. Syst.*, **21** (2003), 315–346.
16. Y. L. Zhu, J. Min, Y. Q. Zhou, Semantic orientation computing based on HowNet, *J. Chin. Inf. Process.*, **20** (2006), 14–20.
17. G. Somprasertsri, P. Lalitrojwong, Mining feature-opinion in online customer reviews for opinion summarization, *J. Univ. Comput. Sci.*, **16** (2010), 938–955.
18. B. Liu, Sentiment analysis and opinion mining, *Synth. Lect. Hum. Lang. Technol.*, **5** (2012), 1–167.
19. B. Pang, L. Lee, S. Vaithyanathan, Thumbs up? sentiment classification using machine learning techniques. In *Proc. Conf. Empir. Methods Nat. Lang. Process.*, Philadelphia, USA, (2002), 79–86.
20. A. Go, R. Bhayani, L. Huang, Twitter sentiment classification using distant supervision, *CS224N Project Report, Stanford*, **1** (2009), 1–12.
21. M. Guerini, L. Gatti, M. Turchi, Sentiment analysis: How to derive prior polarities from SentiWord Net, *arXiv preprint arXiv: 1309.5843*.
22. A. Tripathy, A. Agrawal, S. K. Rath, Classification of sentiment reviews using n-gram machine learning approach, *Expert Syst. Appl.*, **57** (2016), 117–126.
23. Y. KIM, Convolutional neural networks for sentence classification, *arXiv preprint arXiv: 1408.5882*.
24. K. Cho, B. van Merriënboer, D. Bahdanau, On the properties of neural machine translation: encoder-decoder approaches. In *Proce. Eighth Workshop Syntax, Semant. Struct. Stat. Trans.*, Doha, Qatar, (2014), 103–111.
25. Z. Qu, Y. Wang, X. Wang, A Hierarchical attention network sentiment classification algorithm based on transfer learning, *J. Comput. Appl.*, **38** (2018), 3053–3056.
26. J. Pennington, R. Socher, C. D. Manning, GloVe: Global vectors for word representation, *Proc. Int. Conf. Empir. Methods Nat. Lang. Process.*, (2014), 147–168.
27. F. Abid, M. Alam, M. Yasir, C. Li, Sentiment analysis through recurrent variants latterly on convolutional neural network of twitter, *Future Gener. Comput. Syst.*, **95** (2019), 292–308.
28. J. Trofimovich, Comparison of neural network architectures for sentiment analysis of Russian tweets, In *Proc. Int. Conf. Dialogue 2016*, RGGU, (2016).

29. Q. Qian, M. Huang, J. Lei, X. Zhu, Linguistically regularized LSTMs for sentiment classification. In *Proc. 55th Ann. Meet. Assoc. Comput. Ling.*, **1** (2016), 1679–1689.
30. L. Nio, K. Murakami, Japanese sentiment classification using bidirectional long short-term memory recurrent neural network. In *Proc. 24th Ann. Meet. Assoc. Nat. Lang. Process.*, Okayama, Japan, (2018), 1119–1122.
31. M.-L. Zhang, Z.-H. Zhou, A review on multi-label learning algorithms, *IEEE Trans. Knowl. Data Eng.*, **26** (2014), 1819–1837.
32. J. Devlin, M. W. Chang, K. Lee, K. Toutanova, BERT: Pre-Training of deep bidirectional transformers for language understanding, In *Proc. 2019 Conf. North Am. Chapter Assoc. Comput. Linguist. Human Lang. Tech.*, (2019), 4171–4186.
33. S. Seo, C. Kim, H. Kim, et al, Comparative study of deep learning based sentiment classification, *IEEE Access*, **8** (2020), 6861–6875.
34. B. Dzmitry, C. Kyunghyun, B. Yoshua, Neural machine translation by jointly learning to align and translate, In *Proc. Int. Conf. Learn. Representations 2015*, San Diego, CA, USA, (2015), 1–15.
35. D. Li, F. Wei, C. Tan, Adaptive recursive neural network for target-dependent twitter sentiment classification, *Baltimore: Assoc. Comput. Ling.-ACL*, (2014), 49–54.
36. S. M. Grüer-Sinopoli, C. N. Thalemann, Bidirectional recurrent neural networks, *IEEE Trans. Signal Process.*, **11** (1997), 2673–2681
37. S. Zhang, X. Xu, Y. Pang, J. Han, Multi-layer Attention Based CNN for Target-Dependent Sentiment Classification, *Neural Process. Lett.*, **51** (2020), 2089–2103
38. K. Saranya, S. Jayanthi, Onto-based sentiment classification using machine learning techniques. In *Proc. 2017 Int. Conf. Innovations Inf., Embed. Commun. Syst.*, (2017), 1–5.
39. S. Tan, J. Zhang, An empirical study of sentiment analysis for Chinese documents, *Expert Syst. Appl.*, **34** (2008), 2622–2629.
40. W. Li, F. Qi, Sentiment analysis based on multi-channel bidirectional long short term memory network, *J. Chin. Inf. Process.*, **33** (2019), 119–128.
41. Y. Yang, H. Yuan, Y. Wang, Sentiment analysis method for comment text, *J. Nanjing Univ. Sci. Tech.*, **43** (2019), 280–285.



AIMS Press

©2020 the Author(s), licensee AIMS Press. This is an open access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0>)