Emotional Analysis of Jingdong Commodity Review based on Deep Learning

Kang Xu[1, a], Ruichun Hou[1, b], Xiangqian Ding[1, c], Ye Tao[2, d] and Zhifang Xu[3, e]

[1]School of Information Science and Engineering, Ocean University of China, Qingdao 266100, China;

[2]School of Information Science and Technology, Qingdao University of Science and Technology, Qingdao 266061, China;

[3]Qingdao Haier Technology Co. Ltd., Qingdao 266000, China.

[a]1255791952@qq.com, [b]hrch4968@sina.com, [c]Dingxq1995@vip.sina.com, [d]taoye_yn@126.com, [e]X.z.f@foxmail.com

Abstract. Jingdong commodity sentiment analysis aims to find out the e-commerce users' attitudes towards the products and the emotional orientation of the evaluation, helping other users to make correct decisions and effectively shape the quality portrait of the products. This paper proposes a subject-oriented AT-LSTM model method for traditional cyclic neural networks. The model is characterized by combining the Attention mechanism on the traditional LSTM model, highlighting the influence of the key input on the output through the Attention layer, and identifying the subject of the input phrase through the subject recognition algorithm, and then combine the phrase subject and the LSTM hidden layer result through the Attention mechanism. The experimental results show that the proposed method can obtain higher classification accuracy and save the workload of manual labeling.

Keywords: deep learning; commodity review sentiment analysis; recurrent neural network; Attention mechanism; AT-LSTM model.

1. Introduction

With the continuous popularization of e-commerce, the number of online product reviews has risen sharply. Therefore, in the face of such huge unstructured commentary information, merchants and customers cannot accurately grasp product reference information. It is time-consuming to refer to this information by manual browsing and has a certain one-sidedness[1]. Therefore, the use of computers to help users accurately and quickly complete and collate the review information of related products, and emotional analysis of the emotions contained in commodity reviews has become a research hotspot that is increasingly concerned.

The sentiment analysis of commodity reviews refers to excavating and analyzing subjective information such as positions, opinions, and emotions through online product reviews[2,3]. The method of emotional analysis of user product reviews is to classify the sentiment orientation of the text, mainly through language knowledge and established sentiment lexicon; secondly, to classify the features according to the characteristics, mainly using traditional methods such as machine learning to look at them. Make traditional classification and extract features and judge. SVM has achieved good classification results in many classification methods for the emotional analysis of online product reviews. For example, Wang Gang[4] proposed an RS-SVM method based on high-dimensional data features in this problem, based on the Random Subspace in integrated learning, using SVM as the base learner; Wang Wenhua[5] uses the SVM algorithm model, according to the online commodity review. The relationship between the emotional words and the attribute words in the text is used to formulate the corresponding feature classification rules to judge the collocation identification of the two, and to analyze the emotional orientation of the features according to the negative words and the emotional words; Yao Nana[6] combines emotional dictionary with SVM to analyze online product reviews; Pang et al.[7] use three kinds of machine learning algorithms, such as maximum entropy, naive Bayes (NB) and SVM, to conduct sentiment orientation analysis, but only if an emotional dictionary needs to be established.

Copyright © 2019, the Authors. Published by Atlantis Press.
This is an open access article under the CC BY-NC license (http://creativecommons.org/licenses/by-nc/4.0/).

The key part of the above-mentioned emotional orientation analysis method is the feature extraction and the establishment of the emotional dictionary. Therefore, the quantity and quality of the features and the accuracy of the emotional dictionary determine the accuracy of the analysis. However, the establishment of the emotional dictionary is extremely cumbersome and the feature extraction is also time consuming and laborious, and it depends on professional knowledge and experience. In recent years, researchers have continuously explored and explored deep learning and made significant progress in various fields. At the same time, many researchers have made various researches in the field of natural language processing and made major breakthroughs, such as using neural networks to learn text features. , establish language models and other aspects. With the continuous study of deep learning, the deep learning model represented by neural network has become the main model of sentiment analysis. For example, Liang Jun[8] used the deep learning method represented by recurrent neural network to implement emotion classification and proposed the emotional limit transfer model; Gao et al[9] proposed a method of contrast divergence-restricted Boltzmann machine, which is calculated by using the probability size is used to perform sentiment analysis on the text. The deep learning methods proposed above all have good results, but we can also use the method of constructing recurrent neural networks (RNNs)[10,11], that is, to make information persistent through loops. While RNNs have "long-term dependence" problems, LSTM[12] is a special RNN that can solve long-term dependence problems, but LSTM still has no way to judge the emotional polarity transfer, but LSTM still has no way to judge the emotional polarity transfer, and there is no way to judge the importance of the phrase in the input sequence[13].

This paper also uses the method of deep learning. In view of the above-mentioned problem that the LSTM cannot determine which phrases in a sentence play an important role in sentiment analysis, it is necessary to evaluate the importance of each input phrase. The Attention model is introduced to calculate the attention of the input, and the impression of the key input on the output is highlighted to improve the accuracy of the LSTM classification. Since a comment text is composed of multiple phrases, there are multiple subjects and the description of the emotion objects is different, so in order to distinguish the subjects and determine the importance of the input phrase, the word vector of the subject phrase is added in the Attention layer.

2. Related Work

2.1 Deep Learning

In machine learning research, deep learning is a new field. The purpose is to simulate our human brain and establish relevant neural networks to analyze learning. According to the hierarchical simulation structure of our human brain, a similar structure is established and the input data is extracted step by step, so as to establish a mapping relationship between the underlying signals and the high-level semantics. With the deepening of research, deep learning has been successfully applied in the direction of speech recognition and image processing, and more and more deep learning methods are also applied to the NLP direction. The method based on deep learning does not depend on the dictionary, and can automatically extract the text features, so as to better solve the problems caused by the traditional "shallow learning". In 2003, Bengio[14] et al. found a way to build a binary language model using neural networks; Andriy Mnih and Geof-freyHinton proposed a level of thought training language model in 2008[15], and the establishment of deep learning models gradually reached people's The extent of acceptance.

2.2 Recurrent Neural Network

Recurrent neural networks (RNNs) have achieved good results in natural language processing and are widely used. The purpose of RNNs is to process sequence data. In the traditional neural network model, the input layer, the hidden layer, and the output layer are all connected, and the nodes in each layer are not connected to each other. This traditional ordinary neural network is mostly helpless for many problems encountered. The reason that RNNs are called recurrent neural networks is that for a sequence, the current output is also associated with the previous output. That is, the neural network

stores the previous information and applies it to the calculation of the current output, then the nodes in the hidden layer are connected to each other, and the input part of the hidden layer includes not only the input of the input layer but also the previous one. The output of the hidden layer at the moment. To reduce complexity, it is often assumed in practice that the current state is only related to the previous states. The following figure is a typical RNNs:

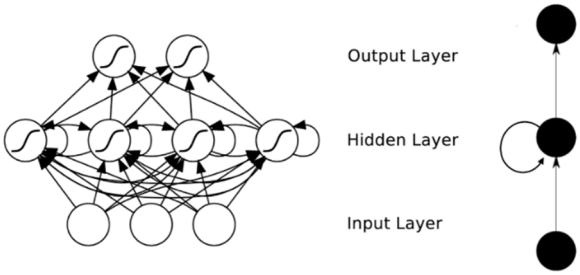

Fig. 1 recurrent neural network

2.3 Emotional Analysis

Sentiment analysis is an emerging research topic. Since being proposed, it has gained a great deal of attention in related fields, especially in the emotional analysis of online commodity reviews, and therefore has a very high research value and application value[16]. As a result, sentiment analysis and related research topics have also received more and more attention in research fields at home and abroad. In 2008, Pang and Lee began to use the word bag model to analyze the sentiment related to the text, followed by more and more people trying to design better engineering models or applying grammatical structure based polarity transfer rules to improve emotional tendencies. The accuracy of the sexual analysis is based on the model proposed by Pang and Lee, and there is no way to obtain deeper semantic information, and the desired effect is not obtained. In 2013, some people used recurrent neural networks to analyze certain data, which is better than traditional methods.In the improved model long-term and short-term memory network (LSTM) of the RNNS, the Attention mechanism is introduced to improve the accuracy of sentiment analysis.

3. Algorithm Model Principle

This section introduces the many drawbacks of traditional recurrent neural networks and leads to its improved model LSTM. However, since LSTM also has an important role in the analysis of sentences in a sentence, it is necessary to evaluate the importance of each input phrase, and introduce the Attention layer to solve the problem; Secondly, there are multiple phrases in a sentence, there are different subjects and the objects described between the subjects are different, so it is necessary to distinguish different subjects. The AT-LSTM model is proposed by adding the word vector of each phrase body and blending the results of the LSTM hidden layer in the Attention layer to calculate the probability of the input.

3.1 Long-term and Short-term Memory Network (LSTM) Model

Recurrent neural networks is a structure that has been added to the loop in traditional neural networks, so the contents of the loop body are executed at each step.At the same time, there are two more serious problems: Firstly, when the input sequence is too long, the influence of the historical node on the circulating neural network is reduced, so that it can not remember the information that is far away from the present, and the role of RNNs in the early stage is not obvious. Secondly, when the sequence entered in the model is too long, a gradient explosion or gradient disappears due to some specific reasons.Thus, an improved model of the recurrent neural networks, LSTM, has emerged. The main change is the introduction of three types of gates at each index position: memory gates, forget gates, and output gates. Through these three door structures, it is used to control and maintain the state information of the unit, store and modify the cell state information, selectively add information

or delete existing information according to the result, and solve the long-term gradient explosion of RNNs. As well as the disappearance of the problem, it also realized long-term memory.

The specific network structure of LSTM and its internal structure are shown in Figure 2:

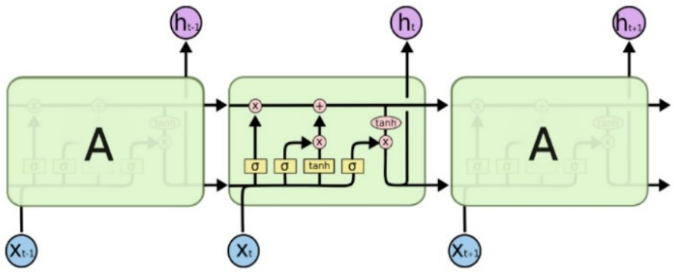

Fig. 2 LSTM model structure

The steps for the LSTM to update the internal state during training are as follows:

In the first step, the LSTM needs to pass the forgetting gate to determine the information that the cell state needs to be discarded. The forgetting gate obtains a value between 0 and 1 according to the historical information of the previous state and the currently input information to represent the probability of forgetting the state of the upper layer of cells. That is to say, the forgetting gate is to control whether or not to forget, that is, to control whether the state of the upper layer hidden cells is forgotten with a certain probability. Using the activation function (generally the sigmoid function), The hidden state of the input previous sequence is denoted as h_{t-1} and the input of this sequence is recorded as x_t by which the output of the forgetting gate between [0.1] is recorded as $f(t)$; 0 represents all information of the discarded history state, and 1 represents all information about the historical status. Where x_t represents the input of t time, h_{t-1} represents the value of the hidden layer at $t-1$ time, W_f, U_f, and b_f represent the weight of the LSTM hidden layer in the forgetting gate, the weight of the current input, and the offset in the forgotten gate; $f(t)$ represents the output of the forgetting gate. The specific calculation is as follows:

$$f(t) = \sigma(W_f \cdot h_{t-1} + U_f \cdot x_t + b_f) \tag{1}$$

In the second step, the LSTM determines the update information through the input gate and places it in the cell. After the circulatory neural network has forgotten part of the previous state, it needs to supplement the latest memory from the current input, and the input gate is responsible for processing the input of the current sequence position to supplement the latest memory. The input gate is composed of two parts. The first part uses the sigmoid function to output as i_t, the second part uses the tanh function to output as a_t, and the two obtain the information updated to the cell state by the bitwise product, where W_i, U_i, W_a, U_a, b_i, b_a are the coefficients and offsets of the linear relationship. The specific calculation is as follows:

$$a_t = \tanh(W_a \cdot h_{t-1} + U_a \cdot x_t + b_a) \tag{2}$$

In the third step, the cell state is updated. After the first two "gates", the deletion and addition of the delivery information can be determined, that is, the update can be performed. Both the results of the Forgetting Gate and the results of the input gate will act on the Cell Status C_t and put the C_{t-1} update to the final cell state C_t, That is:

$$C_t = C_{t-1} \odot f_t + i_t \odot a_t \tag{3}$$

Finally, the result of the final cell output is determined by the output gate. First, the input information is determined by the sigmoid layer to determine the output of the input information, and

then the cell state is processed by the tanh function. Finally, the final output is determined by bitwise multiplication with the previously obtained sigmoid value, and the specific calculation is as follows:

$$O_t = \sigma(W_o \cdot h_{t-1} + U_o \cdot x_t + b_o) \tag{4}$$

$$h_t = O_t \odot f_t + \tanh(C_t) \tag{5}$$

3.2 Attention Mechanism

When people observe things, they tend to selectively focus on some key and decisive information, while ignoring other perceptible secondary information. This mechanism is often called Attention[17]. For example, in people's visual processing, although we can accept the ability of a large field of vision, we can only pay attention to a small part of it; for example, when we read, we can understand the meaning by the core meaning. Therefore, the core idea of the Attention model is to allocate more attention to the part that plays a decisive role, and to allocate less attention to other lesser parts, so that it is more reasonable to use limited computing resources to reduce the impact of secondary factors. At first, the Attention mechanism was widely used in computer vision, such as physical detection[18] and region recognition[19], and achieved good results. The Attention mechanism is just a methodology or mechanism, and there is no strict mathematical definition. In natural language processing, the most frequently used scene for Attention is the Encode-Decode model.

As mentioned earlier, the standard LSTM does not recognize which phrases in the sequence play a decisive role in sentiment analysis, so the importance of the relevant phrases needs to be evaluated. Therefore, LSTM adds the Attention mechanism, which enables LSTM to remember critical information and ignore unimportant information.

3.3 Subject-Oriented AT-LSTM Model

It can be known from the above Attention mechanism that the method for improving the classification accuracy of the LSTM model is to emphasize the influence of the input on the output by calculating the attention of the input. For the improvement of LSTM, the Attention mechanism is added more, but the Attention mechanism is only aimed at the whole sentence. The key information in the sentence is relatively heavy, and there is no careful attention to the subject of each phrase. Since the emotion description objects between the subjects are different, the difference is also large, and the body of each phrase in the sequence is obtained by means of the subject discrimination algorithm that is queried. The AT-LSTM model is proposed by distinguishing different subjects and determining the importance of each phrase to the result. The word vector of the phrase subject is added to the Attention layer and the implicit layer result of the LSTM is combined to calculate the probability distribution of the input.

4. Algorithm Model for Sentiment Analysis

The main design idea of the subject-oriented AT-LSTM network model is to add the word vector of the subject in the Attention layer and combine it with the hidden layer vector of LSTM to form a new combined vector, and then calculate the probability distribution of the input sequence; The hidden layer vector is bitwise multiplied to obtain the representation of Attention, and then the previous LSTM output is nonlinearly combined with the representation of Attention to obtain the final sentence feature representation. Finally, add a linear layer to the feature representation of the sentence, and use the softmax classifier to obtain the final classification result. The specific model structure diagram is shown below:

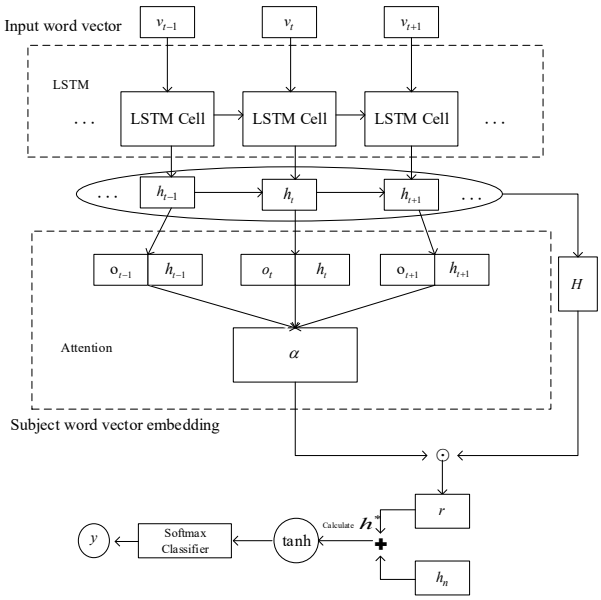

Fig. 3 Body-oriented AT-LSTM model structure

It can be seen from the above specific model framework diagram that the model is mainly composed of three parts: the LSTM part, the input sequence probability distribution of the Attention layer, the obtained sentence feature representation and the final classification result representation. The specific calculation is as follows:

For the LSTM part: text classification based on deep learning, we first train the word vector, and then perform the word vector matrix representation of the comment data. This paper trains the word vector based on the Skip-gram algorithm provided by Googel;

Here, the word vector sequence $V=\{v_1,v_2,v_3,\cdots,v_n\}$ of the previously trained comment text is used, where v_t represents the t-th word vector, and the LSTM model is run for the input sequence to obtain the hidden layer vector $H=\{h_1,h_2,h_3,\cdots,h_n\}$, indicating the length after the relevant sentence segmentation;

Calculate the probability distribution of the input sequence of the Attention layer: First, we need to perform the subject recognition algorithm on the sentence, and get the word vector $O=\{o_1,o_2,o_3,\cdots,o_n\}$ of the subject we need, and the length should be the same as the length of the word vector of the input sequence V. And in order to smoothly combine the subject vector V and the hidden layer vector H, the number of hidden units should be equal to the dimension of the word vector M, and the attention A of the Attention layer is calculated as follows:

$$A=\tanh(W^H\cdot H+W^O\cdot O) \tag{6}$$

Among them, W^H and W^O are projection matrices that need to be trained, and the input sequence probability distribution generated by the Attention layer is α:

$$\alpha=soft\max(W^T\cdot A) \tag{7}$$

W^T is the parameter of SoftMax;

Obtain the special expression of the sentence and output the classification result: first calculate the attention weight by the probability distribution obtained by the Attention layer, expressed as r, and calculate as:

$$r=H\cdot\alpha^T \tag{8}$$

Compared with the simple method of expressing the sentence by LSTM last output $W^x\cdot h_n$, the nonlinear combination of the LSTM output vector and the above-mentioned calculated attention

weight r is better. Finally, we will get the sentence feature representation h^*, which is calculated as:

$$h^* = \tanh(W^p \cdot r + W^x \cdot h_n) \tag{9}$$

Where W^p and W^x are projection matrices that need to be trained, and finally the resulting sentence features are represented by a linear layer, and the softmax classifier is used to obtain the final classification result, which is calculated as:

$$y = soft\max(W^s \cdot r + b^s) \tag{10}$$

Where W^s and b^s are the parameters of the softmax classifier;

5. Experimental Results and Analysis

5.1 Data Set and Data Preprocessing

In order to verify the validity of the above model, the data source of this paper is the 45,000 comments on home appliances (refrigerators) from Jingdong Mall (www.jd.com) from 2016 to 2018. Here, in order to determine the accuracy of the automatic labeling, the comments of four stars and above are regarded as favorable comments, the comments of two stars and below are regarded as bad reviews, and the rest are treated according to the middle evaluation. Because the rating of Jingdong Refrigerator Review is basically above 90%, and finally about 500,000 comments crawled by reptiles (not including the middle review is only praise and bad review), the praise number is about 450,000, and the bad review is about 50,000. Because this paper studies the problem of emotional dichotomy, in order to balance the number of positive and negative samples of training and testing, 22500 comments and 22,500 bad reviews were randomly selected. Of the 45,000 comments, 40,000 were used for model training and 5,000 were for model testing. Finally, the above data set is applied to the training and testing of the model.

The preprocessing of data is also an important step in natural language processing. It is also very important for this experiment. The result of data processing is also directly related to the accuracy of sentiment classification. First, the training of the word vector is required, followed by the word vector matrix representation of the comment data. This paper is based on the Skip-gram algorithm provided by Google to train the word vector, obtain the corresponding Chinese corpus in Wikipedia, and then use the algorithm to train the 50, 100, 200 and 300 dimensional word vectors respectively, and provide them to the AT-LSTM model. Used for sentiment classification and using a word vector filling algorithm to generate a word vector matrix.

5.2 Model Comparison Experiment

Based on the refrigerator review data set provided above, the effectiveness of the subject-oriented AT-LSTM model was verified, and experiments were performed and compared with the experimental classification results based on RNN, LSTM, GRU, and Bi-LSTM. The results of the experiment are shown in the following figure:

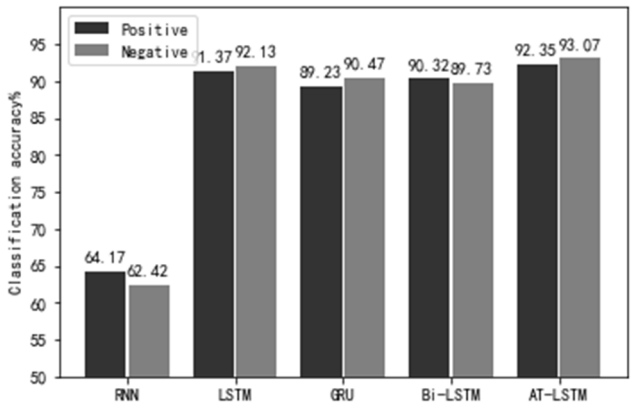

Fig. 4 Comparison of classification accuracy of different models on data sets

It can be seen from the above experimental results:

From the classification accuracy of the above models, it can be seen that the subject-oriented AT-LSTM model is relatively the highest in accuracy compared to the RNN model and other LSTM-based models;

The classification effect of the basic LSTM model is superior to Bi-LSTM and GRU in this data set, while Bi-LSTM and GRU are simplification based on LSTM gate and complication of LSTM structure respectively, and also illustrate the selection of cyclic neural network. At the time, LSTM should be used to improve the addition of the Attention layer instead of the other two models;

5.3 The Influence of Word Vector Dimension on Each Model

In order to investigate the influence of different dimensional word vectors on classification accuracy, this paper compares RNN, LSTM, GRU, Bi-LSTM and subject-oriented AT-LSTM models. The word vector dimensions of the training required for the experiment are 50-dimensional, 100-dimensional, 200-dimensional, and 300-dimensional, respectively. The data set is still the refrigerator review data set, and the experimental results are shown in the following figure. It can be seen from the figure that when the word vector dimension is 100 or more, the accuracy of the classification remains basically unchanged. It is indicated that when the word vector dimension is greater than 100 dimensions, the dimension of the ascension word vector does not have much influence on the improvement of classification accuracy.

The results of the experiment are shown below:

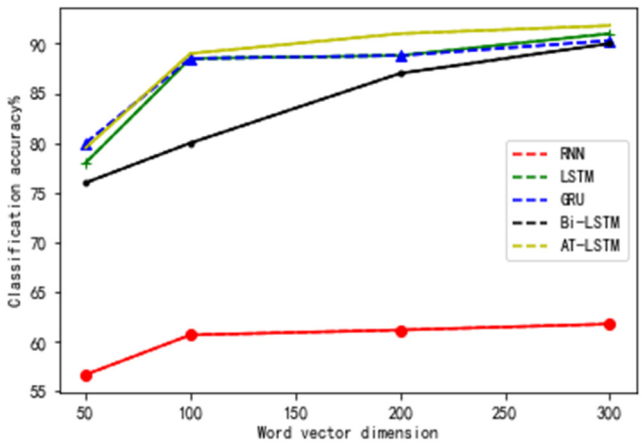

Fig. 5 Comparison of classification accuracy of each model in different word vector dimensions

6. Summary and Future Work

Based on the traditional cyclic neural network, this paper proposes a subject-oriented AT-LSTM sentiment analysis model, which performs sentiment analysis on the comment text data, and compares the model with other cyclic neural network models. Through the traditional analysis of traditional RNNs, and related improvements in its variant model LSTM, the enhanced model's relevance to keywords improves the accuracy of model classification. The experimental results show that compared with the method, the sentiment analysis method proposed in this paper improves the accuracy and verifies the validity of the model.

Because the deep learning method is used to analyze the text emotionally, it is not necessary to establish an emotional dictionary, so it can be used in different fields for sentiment analysis. Therefore, in the subsequent research, the model will be used to analyze the emotions in different fields. Of course, using deep learning methods to study sentiment analysis, there are still many problems that need to be further explored. For example, in this experiment, whether different data sets will cause certain differences in experimental results and the current model in this paper only verifies the emotional classification. The two-category model, but the results of classification on the three-category and five-category problems are not satisfactory. Therefore, in the subsequent work, the model needs to be further optimized, and different data sets are used for training test. Further optimization is needed to find a more appropriate deep learning sentiment analysis algorithm.

Acknowledgements

This work was financially supported by National Key Research and Development Plan 2017YFB400903.

References

[1]. REN Y Y, WANG W P. Extracting product features and determine sentiment orientation from Chinese online reviews[J]. Computer Systems & Applications, 2014,23(11): 22-27.

[2]. Zhang Z Q, Ye Q, Li Y J. A Review of Emotional Analysis of Internet Commodity Reviews[J]. Journal of Management Sciences,2010,13(006):84-96.

[3]. Pang B, Lee L. Opinion mining and sentiment analysis[J]. Foundations and Trends® in Information Retrieval,2008, 2(1-2): 1-135.

[4]. WANG G, YANG S L. Study on Emotional Analysis of Network Commodity Reviews Based on RS-SVM[J]. Computer Science,2013,40(11):274-277.

[5]. WANG W H, ZHU Y H, XU Y Q, et al. Analysis on emotional tendencies of attribute characteristics in product reviews based on SVM[J]. Journal of Hunan University of Technology, 2012, 26(5): 76-80.

[6]. YAO N N. Research on sentimental classification of product reviews based on machine learning[D]. Beijing: Capital Normal University, 2013: 41-46.

[7]. PANG B, LEE L, VAITHYANATHAN S. Thumbs up? Sentiment classification using machine learning techniques[C]//EMNLP 2002: Proceedings of the ACL-02 Conference on Empirical Methods in Natural Language Processing. Stroudsburg, PA, USA: Association for Computational Linguistics,2002, 10: 79-86.

[8]. Liang J, Chai Y M, Yuan H B, et al. Sentiment Analysis Based on Polar Transfer and LSTM Recursive Network[J]. Journal of Chinese Information Processing,2015,29(5):152-159.

[9]. Gao W, Chen B F, et al. Speech analysis of product reviews based on contrast divergence-restricted Boltzmann machine deep learning[J]. Journal of Computer Applications, 2016,36(4):1045-1049.

[10]. Zhu Q X, Sun F. Application Research of RNN Neural Network[J]. Journal of Beijing University of Chemical Technology: Natural Science Edition 1998(1):88-92.

[11]. Taniar D, Safar M, Tran Q T, et al. Spatial network RNN queries in G1S[J]. The Computer joumal,2011 ,54(4) :617-627.

[12]. Hu X C. Study on Semantic Relation Classification Based on LSTM[D]. Harbin: Harbin Institute of Technology, 2015.

[13]. Daniel Soutner, Ludek Muller. Application of LSTM Neural Networks in Language Modelling [M]. Springer Berlin Heidelberg: 2013.

[14]. Y. Bengio, R. Ducharme, P. Vincent, et al. A neural probabilistic language model[J]. Journal of Machine Learning Research,2003,3:1137-1155.

[15]. Mnih A, Hinton G E. A Scalable Hierarchical Distributed Language Model[C]//Proceeding of NIPS.2008:1081-1088.

[16]. Zhao Y Y, Qin B, Liu T. Text Emotion Analysis[J]. Journal of Software,2010,21(8): 1834-1848.DOI:10.3724/SP.J.1001.2010.0-3823.

[17]. Anderson J R. Cognitive psychology and its implications[M]. WH Freeman/Times Books/ Henry Holt & Co, 1985.

[18]. Liu T, Yuan Z, Sun J, et al. Learning to detect a salient object[J]. IEEE Transactions on Pattern analysis and machine intelligence, 2011, 33(2): 353-367.

[19]. Achanta R, Estrada F, Wils P, et al. Salient region detection and segmenation [C]// Intemational conference on computer vision systems. Springer, Berlin, Heidelberg, 2008: 66-75.