Aspect sentiment analysis based on gating convolutional network and attention weighting mechanism

Fan Xu
College of Artificial Intelligence and Computer
JIANGNAN UNIVERSITY
WUXI, China
e-mail: 2631162557@qq.com

Xuezhong Qian
College of Artificial Intelligence and Computer
JIANGNAN UNIVERSITY
WUXI, China
e-mail:qxzvb@163.com

***Abstract*—Aspects Sentiment analysis is a fine-grained text on emotional classification. Aiming at the problem that traditional attention mechanism can't effectively combine contextual meaning an spectoward with information, and single level attention can't obtain deep emotional information features, a gated convolutional network model with attention weights is proposed. Firstly, the word layer is modeled by two-way long-term and short-term memory network, and context semantic information is captured in different directions. In the meantime, different weights are assigned to context words with different positions, and then sentences are gated by convolutional network. The layers are modeled to capture the importance of different sentences, and finally the softmax regression is used for classification. The laboratory finding on the Restaurant DS and the Laptop DS in SemEval2014 indicate that the classification accuracy is better than the classification effect of GCN.**

Keywords-attention-weight;bilstm;gcn;deep learning; Aspect-based analysis;

I. INTRODUCTION

Unlike ordinary sentiment analysis, aspect-based sentiment classification is a fine-grained task in sentiment analysis whose purpose is to extract the sentiment tendency of a specific target in a given text. For instance, the sentence "The cup is easy to use, but the delivery is very slow." [1]The emotional tendency of the target "cup" is positive, while the emotional tendency of the target "courier" is negative. This example involves the evaluation of two aspects of commodities, and the emotional tendency involved in each aspect is different. Therefore, if the overall emotional judgment of the sentence is made, it is not accurate. The sentiment analysis of specific objects is more targeted and more valuable than the overall tendency of the text, so it is also receiving as far as attention from researchers.

Over the years, neural network models have been extensively used in various tasks in the field of NLP, such as sentiment analysis and machine translation [3], because they do not depend on complex feature engineering and can fully mine feature information of texts. And automatic summaries. LSTM can solve the gradient explosion problem, and the effect is better than standard RNN, so it has been widely used in aspect emotion classification. For instance, Wang et al. [4] proposed two LSTM models based on the attention mechanism, namely AE-LSTM and ATAE. -LSTM. The addition of the attention mechanism makes the neural network model perform better feature extraction when training to improve the accuracy of classification

The traditional attention mechanism cannot effectively combine aspect words and contextual word meaning information, and only using the LSTM model with strong modeling ability for sequential features will focus on aspect-related semantic modeling and weaken the overall text semantic Considerations.Based on this, this paper proposes an aspect sentiment analysis model based on gating convolution and attention weighting mechanism. A special attention weighting method is used to strengthen the semantic information, while combining gated convolution and LSTM to simultaneously model aspect-related semantics and overall text semantics. Experimental results prove that the method presented in this article is efficient.

II. ASPECT SENTIMENT ANALYSIS MODEL

A. Task definition

$S=\{W_1,W_2,\cdots\cdots,W_a,\cdots\cdots,W_n\}$ is defined as a sentence of some words,which contains a few of aspect words $\{w_1,w_2,\cdots\cdots,w_m\}$ .The task of aspect sentiment analysis are to judge the emotional tendencies according to the specific meaningful words in the word list. Word vectors can better learn the semantic information contained in words in a low-dimensional space. This paper maps the words in the data set to continuous dense low-dimensional vectors.the whole word data are stored on the word embedding vector $V_w \in R^{d\times|v|}$,d represents the dimension of the word vector,n represents the size of the matrix.

B. Attention weight layer

Generally speaking, word order information has a great influence on aspect emotion species. In order to extract context semantic information of aspect words, the model first uses bidirectional LSTM to obtain emotional information about the front and back of aspect words. In either orientation, $LSTM_L$ deal with the list entries of the aspect word from the left, while $LSTM_R$ deal with the list entries of the opinion word from the right , thus, it creates two indie LSTMs, an output list and a latent state.At the same time, we think using the face word take for latent state can make better of the meaning of the face list, definite a length range of the whole sentence, and also avoid causing the damage of available emotional information .

The word vector matrix of sentence S is

2473-3636/20/$31.00 ©2020 IEEE
DOI 10.1109/DCABES50732.2020.00023

$E_s = \{e_1, e_2, \cdots\cdots, e_{a-1}, e_a, e_{a+1}, \cdots\cdots, e_n\}$, e_a is the word embedding of the aspect word. The LSTM model mainly includes the afferent gate i_t , the forgotten gate f_t ,the input gate o_t ,the memory unit c_t ,the hidden state h_t .The updated status over time is as follows:

$$f_t = \sigma\left(W_f\left[h_{t-1}; e_t\right] + b_f\right) \quad (1)$$

$$o_t = \sigma\left(W_o\left[h_{t-1}; e_t\right] + b_o\right) \quad (2)$$

$$g_t = \tanh\left(W_r\left[h_{t-1}; e_t\right] + b_r\right) \quad (3)$$

$$c_t = i_t \odot g_t + f_t \odot c_{t-1} \quad (4)$$

$$h_t = o_t \odot \tanh(c_t) \quad (5)$$

Two hidden state vector matrices:

$$H = \{h_1, h_2, \cdots\cdots, h_{a-1}, h_a\} \in R^{n*d} \quad (6)$$

$$H^{'} = \{h_n^{'}, \cdots\cdots, h_{a+1}^{'}, h_a^{'}\} \in R^{n*d} \quad (7)$$

represents the relevant meaningful content generated from our model, where n is the extent of a word list， d is the number of conceal layers, h_a and $h_a^{'}$ is the hidden condition of the function word.

We first use the conceal state of the final time node. In order to take advantage of the hidden condition of every time node, an heed weighting machine-made is programed and rooted.According to the different emotional contribution of each word, the attention weight is assigned to different hidden states. Specifically, attention weights are obtained by hiding state and aspect words.

In the model, in order to combine the aspect word information, the sum containing the semantic information h_a and $h_a^{'}$ of the aspect words in different directions is connected into the aspect word sense vector $\left[h_a^T, h_a^{'T}\right]^T$ .By multiplying with the matrix, the closet condition H and H 'are mapped to the meaning world. The attention weight is assigned as follows:

$$\alpha = soft\max\left(H \times W_M \times \begin{bmatrix} h_a \\ h_a^{'} \end{bmatrix}\right) \quad (8)$$

$$\alpha^{'} = soft\max\left(H^{'} \times W_M \times \begin{bmatrix} h_a \\ h_a^{'} \end{bmatrix}\right) \quad (9)$$

W_m is a weighted data calculated by gathering the weighted latent condition in the bidirectional LSTM.

The final hidden state is expressed as follows:

$$\gamma = \alpha^T H \quad (10)$$

$$\gamma^{'} = \alpha^{T'} H^{'} \quad (11)$$

C. Gated convolutional layer

Our model is built on convolutional layers and gating units. Each convolution filter calculates n features at different granularities from the embedding vector at each position. The gating unit on top of the convolutional layer at each location is also independent of each other. In addition, our model is also equipped with two effective filtering mechanisms: the gating unit at the top of the convolutional layer and the largest pool layer, both of which can accurately generate and select emotional features related to the aspect.

Specifically, the input sentence is represented by a matrix through the embedding layer, $X = [v_1, v_2, \cdots\cdots, v_L]$,where L is the extent of the filling word list.Convolution filter $W_C \in R^{D \times K}$ map the k words on the received field to a single feature c.When we slide the filter through the entire sentence, we acquired a series of new features $c = \left[c_1, c_2, \cdots\cdots, c_L\right]$.

If there are filters of the same width, the output features form a matrix. For each convolutional filter, the maximum timeout pool layer takes the maximum value in the generated convolutional features and generates a fixed-size vector equal to the number of filters.Specifically, we will calculate c_i as:

$$a_i = relu(X_{i:i+k} * W_a + V_a v_a + b_a) \quad (12)$$

$$s_i = \tanh(X_{i:i+k} * W_s + b_s) \quad (13)$$

$$c_i = s_i \times a_i \quad (14)$$

D. Sentiment classification layer

The output of the sentence attention layer is input to the emotion classifier layer for classification, and the softmax function is used to calculate the emotion category probability. The calculation process is as follows:

$$P = \mathrm{Softmax}(W_P S + b_p) \quad (15)$$

Among them, and are the weight parameter and bias parameter respectively, and S is the output of the sentence attention layer.

III. Experiment and result analysis

A. Data sets

The experiments in this paper use the Laptop and Restaurant datasets of SemEval2014 Task4. These two data sets contain user's comment information, target objects and emotional tendency labels. The emotional tendency is grouped into three kinds.Maintaining the Integrity of the Specifications

Table 1. Statistics of the experiment datasets

Data sets		positive	negative	neutral
Laptop	Training set	994	464	128
	Test set	341	169	870
Restaurant	Training set	2164	637	807
	Test set	728	196	196

B. Parameter settings

The experiment uses 300-dimensional Glove word vector to initialize the data set. The size of the hidden layer is the same as the dimension of the word embedding, especially the dimension of the concentration weight and the extent of the input word list are equal . The L2 regularization weight is programed to 0.001, the learning rate is set to 0.01, and the batch size is set to 32. The parameter α is initialized with 1, the parameter β is initialized with 0, and the other parameters are distributed uniformly through U (-0.01,0.01).

C. Evaluation index

for the sake of quantize the performance of aspect-level sentiment classification, this paper uses accuracy rate as the evaluation index.

$$Acc = \frac{T}{N} \tag{16}$$

Among them, Acc is the percentage of correctly predicted samples in all specimen, T is the number of properly predicted specimen, and N is the sum of samples. In general, the greater the Acc, the greater the accuracy and the higher the accuracy of the model performance.

D. Contrast model

The comparison model of this article is as follows:

1) CNN. Adopt classic convolutional neural network structure: convolution, pooling, fully connected layer and softmax layer, regardless of aspect information.

2) LSTM. Based on the standard LSTM of the recurrent neural network, the LSTM is used to obtain the hidden state of each word in the sentence, and then the last hidden state in the LSTM is used for emotion classification.

3) TD-LSTM. Two LSTMs are used to model the left object and the right context, respectively, to predict the emotional tendency of the aspect.

4) BiLSTM. BiLSTM contains two standard LSTMs. You can get the output in the forward scan and the other in the reverse scan of the text.

5) ATAE-LSTM. Based on the attention mechanism of LSTM, the hidden state of LSTM is weighted and summed according to the attention weight.

6) GCN. Stands for Gated Convolutional Network, but no aspect information is considered.

E. Result analysis

Table 2 shows the overall classification accuracy of the 7 methods on the Laptop and Restaurant datasets.

Table 2.Classification accuracy of different methods

Methods	Laptop	Restaurant
CNN	33.39	66.33
LSTM	66.50	74.30
TD-LSTM	68.10	75.60
BiLSTM	68.26	76.40
ATAE-LSTM	68.70	77.20
GCN	68.34	79.30
AE-GCN	72.10	79.57

From the results given in Table 2, AE-GCN has achieved good results on both data sets. Because natural language processing has timing characteristics, CNN is usually not used to deal with the problem, so the effect of CNN is also the worst. The memory unit in the LSTM model can effectively record the text timing information, and compare the accuracy of the CNN algorithm. The results have been significantly improved. Most algorithms are also improved based on the LSTM model, but the effect of the LSTM model without considering the object information is still very poor, which also verifies the importance of the attention mechanism.

TD-LSTM considers the important role of information in classification, and the experimental results are 1% to 2% higher than LSTM. The ATAE-LSTM with the attention mechanism added has an accuracy rate of 77.2%, which is 1.6% and 0.8% higher than TD-LSTM and BiLSTM, respectively.

Compared with GCN, the accuracy rate of AE-GCN on the Restaurant dataset is increased by 0.27%, especially on the Laptop dataset, the accuracy is improved by 3.4%, and the effect is significantly improved. One side, AE-GCN uses the DLSTMs construction, who deal with sentences from two orientation, takes advantage of positional semantic content from the front and back contexts, and generates more effective attention. On the other hand, our model is built on convolutional layers and gating units. Each convolution filter calculates n features at different granularities from the embedding vector at each position. The gating units at the top of the convolutional layer at each location are also independent of each other and are more suitable for parallel computing.

IV. IMPACT OF GATING MECHANISM

We use a specific paraphrase sentence as an example to illustrate how the proposed GCN works. In other neural networks, the weights generated by the gate are more difficult to achieve than the attention weights in other neural networks. Note that the weight score is an overall score in the word and vector dimensions, and in our model, there is a gate output. Therefore, we use a filter that is only three words wide to train a small model. Then, for each word, we sum the output of the Relu gate. After normalization, we plot the value of each word in the figure below. Considering the goals of different aspects, the Relu gate will control the output of the Tan gate.

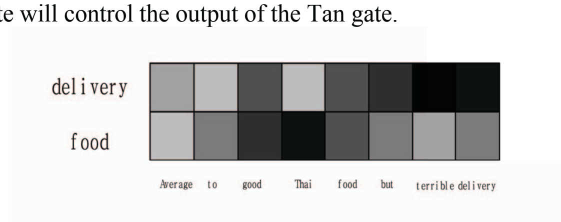

Figure 1.Impact of gating mechanism

V. CONCLUSION

So as to better collect the semantic content of field words and text, this paper proposes an aspect sentiment analysis model based on gating convolution and attention weighting mechanism, which effectively utilizes the correlation between aspect words and context. The input text is modeled by LSTM and GCN, and the object-related semantics are strengthened in the overall semantic scene. The laboratory result reveals that our model has received significant result and is superior to all baseline models. In the future, we will conduct in-depth research on how to apply large-scale sentiment dictionaries to neural networks..

ACKNOWLEDGEMENTS

This work is partially supported by the National Natural Science Foundation of China under Grant No. 61673193 ; the Fundamental Research Funds for the Central Universities of China under Grant Nos. JUSRP51635B, JUSRP51510;

REFERENCES

[1] Jiang, Long, Yu, et al. Target-dependent Twitter sentiment classification[C]// Meeting of the Association for Computational Linguistics: Human Language Technologies. 2012.

[2] ZHANG M S, ZHANG Y, VO D T. Gated Neural Networks for Targeted Sentiment Analysis / / Proc of the 30th AAAI Conference on Artificial Intelligence. Palo Alto, USA: AAAI Press, 2016: 3087-3093.

[3] TANG D, QIN B, LIU T. Aspect Level Sentiment Classification with Deep Memory Network / / Proc of the Conference on Empirical Methods in Natural Language Processing. Stroudsburg, USA: ACL, 2016: 214-224.

[4] YANG M, TU W T, WANG J X, et al. Attention-Based LSTM for Target Dependent Sentiment Classification / / Proc of the 29th AAAI Conference on Artificial Intelligence. Palo Alto, USA: AAAI Press, 2017: 5013-5014.

[5] KIM Y. Convolutional Neural Networks for Sentence Classification / / Proc of the Conference on Empirical Methods in Natural Language Processing. Stroudsburg, USA: ACL, 2014: 1746-1751.

[6] Cho K,Van Merrienboer B,Gulcehre C,et al.Learning phrase representations using RNN encoder-decoder for statistical machine translation[C]//Proceedings of EMNLP 2014,2014.

[7] Wang J,Yu L C,Lai K R,et al.Dimensional sentiment analysis using a regional cnn-lstm model[C]//Proceedings of the 54th Annual Meeting of the Association for

[8] LIN C,HE Y.Joint sentiment/topic model for sentiment analysis[C]//Proceedings of ACM Conference on Information and Knowledge Management.New York,

[9] Srivastava N,Hinton G,Krizhevsky A,et al.Dropout :A simple way to prevent neural networks from Overfitting[J].Journal of Machine Learning Research ,2014,15(1):1929-1958.

[10] Chen T, Xu R, He Y, et al. Improving sentiment analysis via sentence type c1assification using BiLSTM-CRF and CNN[J]. Expert