2020 IEEE 3rd International Conference on Automation, Electronics and Electrical Engineering (AUTEEE) | 978-1-7281-9827-9/20/$31.00 ©2020 IEEE | DOI: 10.1109/AUTEEE50969.2020.9315668

Sentiment Infomation based Model For Chinese text Sentiment Analysis

Gen Li[12],QiuSheng Zheng[12],Long Zhang[12],SuZhou Guo[12],LiYue Niu[12]
[1]Henan Key Laboratory on Public Opinion Intelligent Analysis, ZhengZhou, China.
[2]School of Computer Science, Zhongyuan University Of Technology, ZhengZhou, China.
liamao1995@163.com, {zqs,zhanglong}@zut.edu.cn, {1154369693,474433045}@qq.com

***Abstract*—As an important task of natural language processing,chinese text sentiment analysis aims to analyze the comprehensive sentiment polarity of chinese text.With the emergence of various deep neural network models,sentiment analysis tasks have once again made significant progress.However,these neural network models could not accurately capture sentiment information on sentiment analysis tasks,which leads to their instability.In order to enable the model to explicitly learn the sentiment knowledge in chinese text,this paper proposes a sentiment information based network model(SINM).We use Transfomer encoder and LSTM as model components.With the help of Chinese emotional dictionary,we can automatically find sentiment knowledge in chinese text.In SINM,we designed a hybrid task learning method to learn valuable emotional expressions and predict sentiment tendencies.First of all,SINM needs to learn the sentiment knowledge in the text.Under the auxiliary influence of emotional information,SINM will pay more attention to sentiment information rather than useless information.Experiments on the dataset of ChnSentiCorp and ChnFoodReviews have found that SINM can achieve better performance and generalization ability than most existing methods.**

***Index Terms*—sentiment analysis, sentiment information, neural network**

I. Introduction

With the vigorous development of various Internet technologies, more and more electronic business platform,social platforms,news platforms, film and television platforms have emerged.People only need to browse other people's comments on commodities or movies online as a reference to filter out the commodities they want to buy or the movies they want to watch. Whether it is communication, shopping, movie watching, study, work, etc., users just need to take advantage of the Internet to obtain a large amount of relevant information and post their opinions, opinions, emotions, attitudes online,therefore while people enjoy this lifestyle, innumerable opinions with user's emotion began to appear and made a huge impact on events, goods, services, topics, organizations, etc.As the platform side, they collect user comments or evaluations for analyzing user's sentiment and satisfaction.For instance, news platforms want to analyze the opinions and sentiments of netizens,electronic business platform platforms hope to get user reviews to improve products.With so many demands, sentiment analysis has become a very active field.

As a very interesting and important field in NLP.Internet texts's sentiment analysis has been extensively studied.In 2012, Wei Wei et al. summarized a survey on Chinese texts sentiment analysis, they focus on the traditional methods and key technologies in this field at that time [1].Now, Sentiment analysis generally includes processes such as text feature processing,user opinion extraction,and sentiment opinion analysis.Existing research can be divided into traditional analysis methods relied on sentiment lexicons and sentiment analysis methods based on machine learning, these two methods have their own characteristics, traditional analysis methods have advantages of logical intelligence and interpretability, and machine learning method is supported by data and statistics, In the era of big data, machine learning methods have showed their glorious performance,it usually includes supervised methods and unsupervised methods.In past years,sentiment prediction generally adopt statistical methods for text feature processing and convert text to vector,supervised methods use sentiment polarity as classification labels, and employ some machine learning methods (SVM, Naive Bayes. Maximum entropy, etc.) to train sentiment analysis model [2] [3] [4].The unsupervised method is based on sentiment lexicons, syntactic analysis, etc. to perform sentiment analysis on the text [5] [6].

At present , there are two very popular research direction of sentiment analysis, one is the traditional sentiment analysis method based on sentiment dictionary, and another is the sentiment analysis method based on machine learning or deep learning .

In the field of sentiment dictionary based sentiment analysis,Taboada et al. proposed lexicon-based methods of sentiment analysis in 2011,they prepare a set of positive and negative paradigm words,then they calculate the sentiment polarity of a new word by point mutual information (PMI) and latent semantic analysis (LSA) [7].Saif et al. presented SentiCircles, a lexicon based method of sentiment analysis on Twitter which takes into account the co-occurrence patterns of words with different contexts of tweets to capture their semantics and update their pre-assigned strength and polarity [8].Y Li et al. extracted sentiment words from the buzzwords of HowNet, NTUSD and Sina Weibo posts as seed words to build an sentiment dictionary, then converted the text into vectors with sentiment features, and finally classified the microblogs text using machine learning methods [9].

Recently, under the nourishment of deep learning technology, various network models of sentiment analysis have appeared one after another.Inspired by computer vision,TextCNN and CharSCNN adopt convolutional neural networks to extract

Shenyang, China
November 20-22, 2020

text features and classify text [10] [11].Who and others used LSTM to carry out sentiment classification research on Twitter [12].These papers have made great contributions.Subsequently,considering that CNN is more inclined to extract local features of text, RNN-based networks are time-dependent, and combining their advantages and disadvantages, some methods that combine CNN and RNN network structures have been proposed, and they also show good performance [13] [14] [15].Because these basic models are difficult to pay attention to key information on different NLP tasks, Attention mechanism and Self Attention were proposed [16] [17]. Unfortunately, in the sentiment analysis task, some deep neural network with the attention mechanism is also implicitly learning sentiment information,such as attention based GRU network,attention based LSTM network and Transformer Encoder based network [18] [19] [20] [21] [22], which could not obtain stable knowledge and understanding of the sentiment information in the text.

Dictionary-based sentiment analysis methods have good logic and stability, while deep learning-based sentiment analysis methods can show better performance driven by big data. In order to take into account the respective advantages of the two methods, in this paper, to integrate the sentiment knowledge of sentiment dictionary into the deep neural network model, we designed a multi-task learning method. At the same time, we combined the advantages of Transformer and bidirectional LSTM(BiLSTM) neural network, and append attention mechanism for BiLSTM network.In order to enable SINM to explicitly focus on the distribution of sentiment knowledge in the text for sentiment analysis tasks, we designed a hybrid task Learning method, SINM can have high F1 score and generalization ability in Chinese text sentiment analysis tasks.

II. Sentiment Information Based Model

We will introduce the sentiment information based model in this part, as shown in Figure 1, SINM includes two learning tasks: one auxiliary task and one main tasks.The auxiliary task is the sentiment information recognition task by self-supervised learning , and the main task are the sentence comprehensive sentiment polarity prediction task by supervised learning . They both use the outputs of Transformer encoder as input and dive into different networks of training for their respective tasks . The first subsection below introduced the word embedding layer , the second subsection will explain the Transformer's encoder layer , the third and fourth subsection will introduce the emotional information recognition layer and the bidirectional BiLSTM network of the attention mechanism.

It should be noted that SIBM utilizes a self-supervised method based on the emotional dictionary to learn sentiment information recognition task, and the target word recognition of emotional words is to find nouns, pronouns, entity words, etc. which is near the emotional words as the target words. In this paper, the two positions before and after the emotional word (that is, the window is 2) are selected as the target word discovery range.

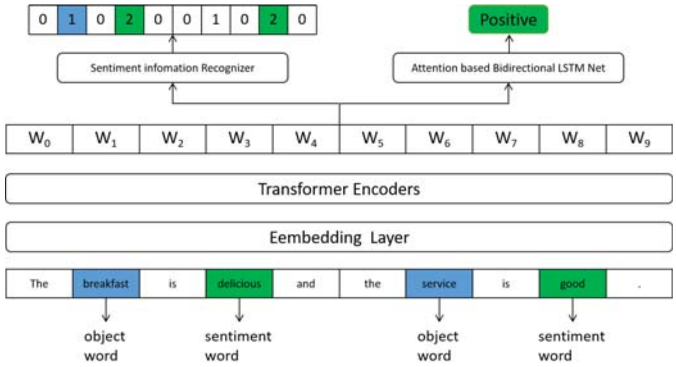

Fig. 1: SINM's Architecture

A. Eembedding layer

First of all, in the part of text preprocessing, we utilize Baidu's open source word segmentation tool LAC to segment the input text [23]. In order to make each word of the input sequence express richer semantic information, we put input sequece to the Embedding layer.Suppose we set d_{model} as the feature dimension of the data in the model,through Embedding layer,The one-dimentional word index sequence x is transformed to word vector sequence with a dimension of d_{model}. The word vectors are coded by meaning coding and position coding . The function of the Eembedding layer is as follows:

$$Embed(x) = [E_{x_1}, E_{x_2}, ...E_{x_{len}}] + \omega_{pos} \tag{1}$$

As Eq(1), The variable matrix $x \in \mathbb{R}^{len \times 1}$, len is the length of input sequence x, x_i represents the value of the i-th value of x,$E \in \mathbb{R}^{v \times d_{model}}$ which is a word vector matrix,v is the size of the vocabulary, E_i represents the i-th word vectors of matric E, $\omega_{pos} \in \mathbb{R}^{len \times d_{model}}$ is the positional parameter.Finaly the word vector matrix E and the positional parameter ω_{pos} can be learned when training the model.

B. Transformer encoder

The classic methods of word representation learning are Word2Vec and Glove [24] [25]. Word2Vec uses the words around the target word to predict the target word as a task. It uses a two-layer neural network for training, and the obtained network weight is used as the word vector. The word vectors learned by word2Vec demonstrated powerful ability in computing similarity between two words in vector space. And GloVe both considered global statistical information and local context information for learning better word vectors. However,in processing the sequence problem,Word2Vec and Glove could not learn the relationship between the input sequence and accurately capture the features required by the downstream task.Transfomer Encoder [17], as a new word representation learning network structure,which could learn richer representation of input data by self attention method.As shown in Figure 2, the word embedding of the input sequence needs to go through a multi-head attention layer and a Feed

Forward layer. Inspired by ResNet, after each layer, a residual connection and layer normalization operation are appened [26]. The multi-head attention layer is based on self attention method, which refers to the attention mechanism, so that each word of the input sequence can learn the degree of association with other words by self attention. The FeedForwad layer is a two-layer neural network with ReLU as the activation function to increase the learning ability of word vectors [27]. After the input sequence passes through the Encoder, it will simultaneously enter to different channels:sentiment information recognizer and the comprehensive sentiment polarity prediction network.The tasks of these two downstream channels will affect their common upstream data.Therefor,the Encoder part can learn more meaningful word vectors for sentiment analysis tasks.Each Encoder is composed of N identical layers, where N is a hyper parameter that can be customized,each same layer is composed of two sub-layers: Multi-Head Attention layer and Feed Forward layer. The attention calculation method in the

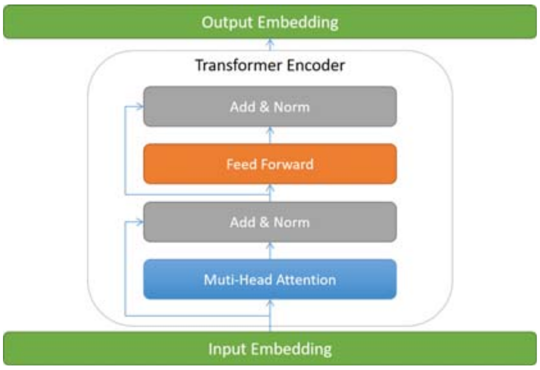

Fig. 2: The Architecture of Encoder

multi-head attention layer is as follows:

$$MutiHead(Q, K, V) = Concat(head_1, ..., head_i)W^O$$

$$\text{where } Head_i = Attention(QW_i^Q, KW_i^K, VW_i^V)$$

where $W_i^Q \in \mathbb{R}^{d_{model} \times d_k}$, $W_i^K \in \mathbb{R}^{d_{model} \times d_k}$, $W_i^V \in \mathbb{R}^{d_{model} \times d_v}$, $W_i^O \in \mathbb{R}^{hd_v \times d_{model}}$,$d_k$,$d_v$ are user-defined parameters and represent the output's feature dimension of W_i^K and W_i^V,h is the number of headers in Multi-Head Attention,the $Concat$ function is concatenating the heads,

$$Attention(Q, K, V) = Softmax(\frac{QK^T}{\sqrt{d_k}})V \qquad (2)$$

Eq(2) is the function of self attention,$\sqrt{d_k}$ is a scaling factor to reduce this influence of larger d_k on the gradient disappearance of the Softmax function.Softmax is a normalized exponential function used to calculate the attention weight of each word with other words.

C. Sentiment Information Recognizer

General sentiment analysis approaches could only implicitly learn the sentiment features in the input sequence in the process of learning parameters,they often could not learn exactly which part of the input sequence will have a strong or weak impact on the sentiment polarity.Therefore, we designed a sentiment information recognizer,the input sequence vector generated by Encoder will pass through the sentiment information recognition part.As Figure 1 shows, the assignment of recognizer is to recognize the type of different word vectors, we have three types of word vectors here: weak-related word types, sentiment word types, and sentiment target word types. The sentiment information recognizer consists of a simple neural network, and the output layer of the neural network has three neurons to recognize different word types.With the sentiment information recognition task training, the recognizer could explicitly capture the distribution of sentiment information in the input sequence, and the word vector of the input sequence generated by the Encoder will also learn stronger sentiment features. For instance, ”The restaurant's breakfast is good”, the word vector of'good' will have more direct sentiment features, the word vector of'breakfast' will have sentiment target features, and other word vectors will obtain weaker sentiment features.

$$SI - Recognizer(X) = W_{SIR}X$$

where $W_{SIR} \in \mathbb{R}^{d_x \times 3}$, d_x is the number of feature dimention of input X.

D. Bidirectional LSTM Net

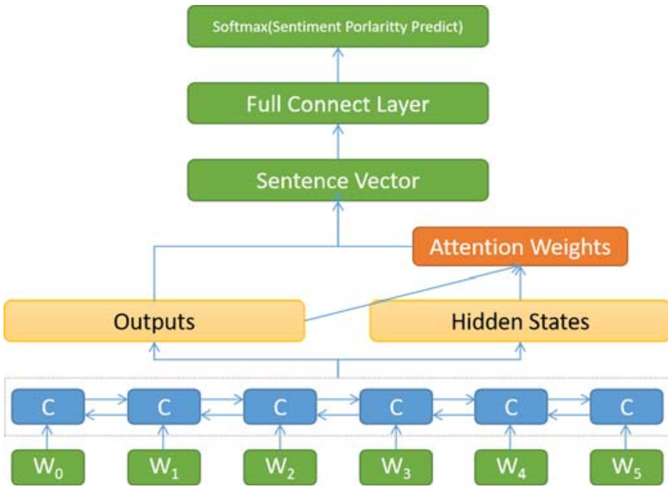

Fig. 3: BiLSTM Net Architecture

Since the input sequence in the sentiment analysis task has temporal characteristics, we adopted the temporal structure based BiLSTM nerwork [28]. Under the sentiment infomation recognition task learning, the word vector sequence generated by Encoder has learned a good distribution of sentiment representation,then We feed the word vector sequence into the BiLSTM network, and at the same time, in order to make the BiLSTM network pay attention to the emotion of these word vectors Feature distribution,in addition, we append attention mechanism to the BiLSTM network [16], Via the BiLSTM network and attention mechanism, the sentiment ralated information of word vector sequence can be understood by BiLSTM net,and the word vector sequence will be transformed into a sentence vector with the influence of sentiment information.In the end, by a layer of neural network and softmax function, the sentence vector will converted into a vector of sentiment tendency for calculating the final sentiment tendency.

For the t-th word vector of the input sequence, LSTM will output its corresponding hidden state h_t [29], the calculation process can be defined as the following:

$$i_t = \sigma\left(W_{ii}x_t + b_{ii} + W_{hi}h_{t-1} + b_{hi}\right)$$

$$f_t = \sigma\left(W_{if}x_t + b_{if} + W_{hf}h_{t-1} + b_{hf}\right)$$

$$g_t = \tanh\left(W_{ig}x_t + b_{ig} + W_{hg}h_{t-1} + b_{hg}\right)$$

$$o_t = \sigma\left(W_{io}x_t + b_{io} + W_{ho}h_{t-1} + b_{ho}\right)$$

$$c_t = f_t \odot c_{t-1} + i_t \odot g_t$$

$$h_t = o_t \odot \tanh\left(c_t\right)$$

where x_t, h_{t-1} and c_{t-1} are inputs at t-th time step, c_t and h_t are the outputs.i_t, f_t, g_t, o_t are input gate, forget gate, cell gate and output gate respectively. h_t and c_t are hidden state and cell state respectively.σ is the sigmoid function.The parameter matrix $W_p \in \mathbb{R}^{d_{hidden} \times d_x}$,$b_p \in \mathbb{R}^{d_{hidden}}$,for $p \in \{ii, if, ig, io\}$.$W_q \in \mathbb{R}^{d_h \times d_{hidden}}$,$b_q \in \mathbb{R}^{d_h}$,for $q \in \{hi, hf, hg, ho\}$.$hidden$ is the dimension of hidden states of LSTM which is customized.x is the dimension of x_t.

Utilizing attention mechanism,see Figure 3 the sentence vector is derived by the following:

$$W_{atten} = Output_{LSTM} H_{len}$$

$$V_{sentence} = W_{atten} Output_{LSTM}$$

where $Output_{LSTM} \in \mathbb{R}^{len \times d_{hidden}}$ which is the hidden states of all time step, $H_{len} \in \mathbb{R}^{d_{hidden} \times 1}$ which is the hidden state of len-th time step.

III. Experiment

A. Dataset and Experiment Setup

To verify the performance of our model on sentiment analysis tasks,we use two Chinese datasets: ChnSentiCorp(https://github.com/SophonPlus/ChineseNlpCorpus/tree/master/datasets/ChnSentiCorp_htl_all) and ChnFoodReviews(https://github.com/SophonPlus/ChineseNlpCorpus/blob/master/datasets/waimai_10k),ChnSentiCorp is a Chinese hotel review dataset which is a unbalanced dataset,ChnFoodReviews is a two-category balanced data set for takeaway-food reviews.Table I presents the detail of datasets,Pos,Neg and Neu are the number of positive,negative,neutral samples.Train and Test are the number of train dataset and test dataset we set.

Table II shows hyper parameter settings of our model in this experiment.

TABLE I: Dataset overview.

Dataset	Pos	Neg	Neu	Train	Test
ChnSentiCorp	5290	2428	48	7266	500
ChnFoodReviews	3000	3000	0	5500	500

TABLE II: Model Setting

Model Parameter	Parameter Description	Value
d_{model}	dimension of data in Encoder	128
d_{hidden}	dimension of hidden state of LSTM	128
N	the number of encoder	1
H	the number of header of per MultipleHead Attention Layer	4
lr	learning rate	0.0001

B. Results and Discussion

In this sentiment analysis task, we use the F1 score of the model in the validation set as the evaluation standard. We respectively used several common models and set them to the same parameters as Table II, and compared with SINM to prove the effectiveness of our model. Table IV shows the performance comparison results of each model. SINM(single task) is the SINM with only one sentiment polarity prediction task. Our model achieved the best results on both data sets. On the Chn-Senti-Corp dataset, SINM's F1 Score is close to 5 percentage points higher than other models.

SINM has two tasks in training process: Task 1: sentiment information recognition task, Task 2: comprehensive sentiment polarity predict task.In this experiment,SINM trained Task 1 and Task 2 together in the first 5 epoches.After the performance of Task 1 is stable, the model starts training only Task 2.

Figure 4 shows the loss function of SINM in training process. It can be found that the model is training two tasks at the same time in first 5 epoches, and the loss function shows a similar downward trend. It is worth noting that after 5 epoches, SINM began to focus on training sentiment polarity predict task, and only on the 8th epoch, the model has reached the best F1 score, which shows that The sentiment distribution learned by Task 1 has made positive effects on Task 2, and it can be seen from Figure 5 that the F1-score fluctuation of the Task 1 after the training is stopped is very small, which indicates that the process of training task 2 alone has hardly effect on Task 1. In brief,auxiliary tasks play a crucial part in promoting the main task.After the 8th epoch, the model gradually entered the over-fitting state, the loss value of the model increased, and the F1 score decreased.

In order to describe the function of SINM more clearly, we visualized the emotional distribution of some text examples. As shown in the Table III, the colored text in the text expresses the sentiment information part of the text recognized by SINM, the text in brackets represents a word of Chinese. The blue font indicates that SINM has recognized the sentiment target word, and the green and red respectively represent the positive and negative sentiment words.While SINM can capture these emotional knowledge well, it means that the mathematical vector of text should have corresponded sentiment features in

TABLE III: Examples of sentiment information distribution of datasets in SINM

Sentiment information distribution of text	Sentiment polarity
Business big bed rooms, the room is (very big), the bed is 2 meters wide, the overall feeling: affordable and good!	Positive
The breakfast is (just so so), the room is clean, but the noise is (too loud), and the price is (relatively expensive)!	Negative
Good, delicious, and inexpensive, fast delivery!	Positive
Delivery time is too long, noodles are soft, and (not chewy)!	Negative

TABLE IV: Best F1 score on different architectures.

F1 score(%) Dataset / Model	ChnSentiCorp	ChnFoodReviews
Atten-BiLSTM [20]	85.40	88.20
Atten-BiGRU [19]	85.20	85.00
CNN-AttenLSTM [21]	85.20	83.2
Transformer-based [22]	83.0977	81.19
Text-CNN [10]	84.4	86.8
SINM(single task)	85.20	86.40
SINM	89.4	89.2

corresponding dimensions. Finally, Under the help of sentiment knowledge, SINM's performance on predicting chinese text's emotional polarity has been greatly improved.

In addition, in order to show the advantages of the multi-task learning model SINM based on emotional information, we also compared the F1 score changes of several different models on the hotel review data set. It can be intuitively seen in Table 5,the performance of SINM is significantly better than other models, and it can quickly reach the best point. In particular, the effect of the SINM(single task) of a single sentiment polarity prediction task is not as good as SINM.

IV. Conclusion

In this article, we propose a deep neural network model based on sentiment information, and design a learning method that mixes main and auxiliary tasks. The auxiliary task based on sentiment words is used to learn the sentiment features and distribution in the text, then the main task could predict the sentiment polarity with more sentiment knowledge. SINM's network model and training method are not complicated, and can achieve excellent results on Chinese text sentiment analysis tasks.

In the future, we prepare to measure more sentiment analysis tasks with SINM.We are also preparing to study how to integrate the natural language features(grammar, part of speech, sentiment punctuation, and sentiment expression, etc.) into feature vectors of text, if the word vectors are equipped with such more sentiment features, it would present better robustness and excellent generalization ability on sentiment analysis tasks.

Acknowledgment

Grateful acknowledgment is made to my supervisor Mr.Zheng, Mr.Zhao and Mr.Zhang who gave me considerable help of suggestion and comments.We also would like to acknowledge the Henan Key Laboratory on Public Opinion Intelligent Analysis for the data and equipment support.

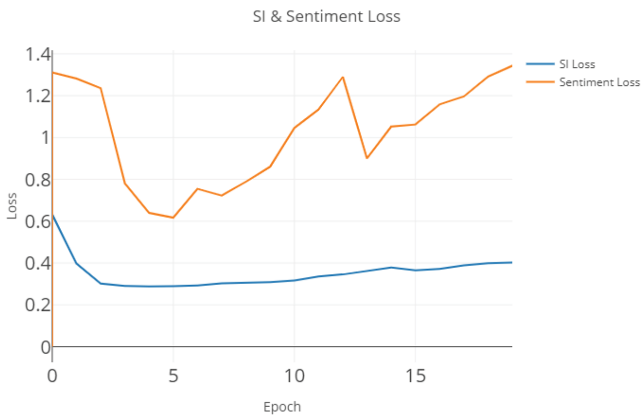

Fig. 4: The loss functions on test dataset of Chn-Senti-Corp.SI Loss is the loss of sentiment information recognition task,Sentiment Loss is the loss of comprehensive sentiment polarity predict task.

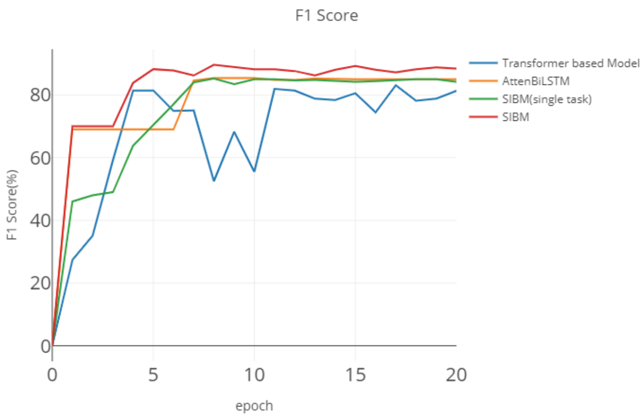

Fig. 5: The F1 score on test dataset of Chn-Senti-Corp

References

[1] W. Wei, Y. Xiang, and Q. Chen, "Survey on chinese text sentiment analysis," *Jisuanji Yingyong/ Journal of Computer Applications*, vol. 31, no. 12, pp. 3321–3323, 2011.

[2] S. Tong and D. Koller, "Support vector machine active learning with applications to text classification," *Journal of machine learning research*, vol. 2, no. Nov, pp. 45–66, 2001.

[3] A. McCallum, K. Nigam *et al.*, "A comparison of event models for naive bayes text classification," in *AAAI-98 workshop on learning for text categorization*, vol. 752, no. 1. Citeseer, 1998, pp. 41–48.

[4] K. Nigam, J. Lafferty, and A. McCallum, "Using maximum entropy for text classification," in *IJCAI-99 workshop on machine learning for information filtering*, vol. 1, no. 1. Stockholom, Sweden, 1999, pp. 61–67.

[5] B. Liu, "Sentiment analysis and opinion mining," *Synthesis lectures on human language technologies*, vol. 5, no. 1, pp. 1–167, 2012.
[6] ——, *Sentiment analysis: Mining opinions, sentiments, and emotions*. Cambridge University Press, 2015.
[7] M. Taboada, J. Brooke, M. Tofiloski, K. Voll, and M. Stede, "Lexicon-based methods for sentiment analysis," *Computational linguistics*, vol. 37, no. 2, pp. 267–307, 2011.
[8] H. Saif, Y. He, M. Fernandez, and H. Alani, "Contextual semantics for sentiment analysis of twitter," *Information Processing & Management*, vol. 52, no. 1, pp. 5–19, 2016.
[9] Y. Li, X. Li, F. Li, and X. Zhang, "A lexicon-based multi-class semantic orientation analysis for microblogs," in *Asia-Pacific Web Conference*. Springer, 2014, pp. 81–92.
[10] Y. Kim, "Convolutional neural networks for sentence classification," *arXiv preprint arXiv:1408.5882*, 2014.
[11] C. Dos Santos and M. Gatti, "Deep convolutional neural networks for sentiment analysis of short texts," in *Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers*, 2014, pp. 69–78.
[12] X. Wang, Y. Liu, C.-J. Sun, B. Wang, and X. Wang, "Predicting polarities of tweets by composing word embeddings with long short-term memory," in *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, 2015, pp. 1343–1353.
[13] C. Zhou, C. Sun, Z. Liu, and F. Lau, "A c-lstm neural network for text classification," *arXiv preprint arXiv:1511.08630*, 2015.
[14] J. Y. Lee and F. Dernoncourt, "Sequential short-text classification with recurrent and convolutional neural networks," *arXiv preprint arXiv:1603.03827*, 2016.
[15] S. Lai, L. Xu, K. Liu, and J. Zhao, "Recurrent convolutional neural networks for text classification," in *Twenty-ninth AAAI conference on artificial intelligence*, 2015.
[16] D. Bahdanau, K. Cho, and Y. Bengio, "Neural machine translation by jointly learning to align and translate," *arXiv preprint arXiv:1409.0473*, 2014.
[17] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," in *Advances in neural information processing systems*, 2017, pp. 5998–6008.
[18] L. Zhou and X. Bian, "Improved text sentiment classification method based on bigru-attention," in *Journal of Physics: Conference Series*, vol. 1345, no. 3. IOP Publishing, 2019, p. 032097.
[19] Y. Pan and M. Liang, "Chinese text sentiment analysis based on bi-gru and self-attention," in *2020 IEEE 4th Information Technology, Networking, Electronic and Automation Control Conference (ITNEC)*, vol. 1. IEEE, 2020, pp. 1983–1988.
[20] P. Zhou, W. Shi, J. Tian, Z. Qi, B. Li, H. Hao, and B. Xu, "Attention-based bidirectional long short-term memory networks for relation classification," in *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, 2016, pp. 207–212.
[21] G. Liu and J. Guo, "Bidirectional lstm with attention mechanism and convolutional layer for text classification," *Neurocomputing*, vol. 337, pp. 325–338, 2019.
[22] A. E. Yüksel, Y. A. Türkmen, A. Özgür, and B. Altınel, "Turkish tweet classification with transformer encoder," in *Proceedings of the International Conference on Recent Advances in Natural Language Processing (RANLP 2019)*, 2019, pp. 1380–1387.
[23] Z. Jiao, S. Sun, and K. Sun, "Chinese lexical analysis with deep bi-gru-crf network," *arXiv preprint arXiv:1807.01882*, 2018. [Online]. Available: https://arxiv.org/abs/1807.01882
[24] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient estimation of word representations in vector space," *arXiv preprint arXiv:1301.3781*, 2013.
[25] J. Pennington, R. Socher, and C. D. Manning, "Glove: Global vectors for word representation," in *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, 2014, pp. 1532–1543.
[26] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.
[27] X. Glorot, A. Bordes, and Y. Bengio, "Deep sparse rectifier neural networks," in *Proceedings of the fourteenth international conference on artificial intelligence and statistics*, 2011, pp. 315–323.
[28] M. Schuster and K. K. Paliwal, "Bidirectional recurrent neural networks," *IEEE transactions on Signal Processing*, vol. 45, no. 11, pp. 2673–2681, 1997.
[29] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural computation*, vol. 9, no. 8, pp. 1735–1780, 1997.