



FuncSA: Function Words-Guided Sentiment-Aware Attention for Chinese Sentiment Analysis

Jiajia Wang¹, Hongying Zan^{1,2(✉)}, Yingjie Han¹, and Juan Cao³

¹ Zhengzhou University, Henan 450001, China
{iehyzan,ieyjhan}@zzu.edu.cn

² Peng Cheng Laboratory, Guangdong 518000, China

³ University of Chinese Academy of Sciences, Beijing 100000, China
caojuan@ict.ac.cn

Abstract. Sentiment analysis is an important natural language processing application that empowers many other technologies, including product review analysis and recommendation systems. Knowledge has been proven crucial for providing supervision information and improving performance. However, Chinese function words' knowledge, especially for degree adverbs, negative adverbs, and conjunctions, which may play an essential role in describing the sentiment polarity, is not well investigated in current Chinese sentiment analysis approaches. In this paper, we propose a **F**unction words-guided **S**entiment-aware **A**ttention model (FuncSA) for Chinese sentiment analysis to leverage function words' knowledge. Specifically, we integrate discrete sentiment lexical information using degree adverbs, negative adverbs, and conjunctions in the Chinese Function Word Usage Knowledge Base(CFKB), and improve self-attention to integrate function words' knowledge into the model. We implement our approach on several open datasets and show that function words are essential in guiding sentiment identification.

Keywords: Chinese text sentiment analysis · Function words · Attention mechanism

1 Introduction

Sentiment analysis refers to extracting, analyzing, understanding, and generating subjective information in natural language. It has been widely applied in many applications, such as public opinion analysis, recommendation systems, review analysis and generation, and business decision-making. External knowledge is often introduced to help classify the sentiment polarity according to the sentiment words such as happy, sad, and fear. However, sentiment lexicons only construct the relationship between words and sentiments and fail to model the relationship between words and phrases. In particular, Chinese is composed of basic character units, which are influenced by the characteristics of the language

itself with complex grammatical structures, diversified semantics, and diversified expressions. Especially considering that function words, a vital means of expression in Chinese, can help express the semantic relationship between content words. Among them, adverbs and conjunctions significantly influence the sentiment and polarity of sentiment words.

Adverbs, especially degree and negative adverbs, are the influencing factors and judgment conditions considered widely in the existing studies on sentiment classifications [1–8]. When they modify sentiment words, the orientation and intensity of the sentiment polarity will change to a certain extent. For example, "unclean/不干净" is composed of the positive adjective "clean/干净" and the negative adverb "not/不", so it has a derogatory meaning. The phrase "too clean/过于干净" and "relatively clean/相当干净" under the modification of the degree adverb "too/过于" and "relatively/相当" express different levels of "clean/干净". Besides, conjunctions are function words with the function of connecting, which can connect words, phrases, clauses, or sentences and indicate causal, inference, hypothesis, conditional, and other linguistic relations [13, 14, 16].

Therefore, this paper proposes a framework, **Function** words-guided **Sentiment-aware Attention** (FuncSA) for Chinese sentiment analysis. FuncSA focuses on the sentiment expressed in the input text and obtains the internal correlated features of sentiment words, adverbs, and conjunctions. In particular, based on calculating the polarity values of emotive words modified by adverbs, different weights are given according to the influence of conjunctions in clauses. The sentiment-aware attention will then adjust the contribution of the sentiment of different clauses. These obtained sentiment semantic representations are effectively utilized to guide the model in conducting Chinese sentiment analysis.

2 Related Work

2.1 Sentiment Analysis

Function Words Irrelevant Methods. Neural networks with attention mechanisms can increase interpretability and sentiment analysis performance [9–12]. Cheng [10] makes the model learn more precise and abundant information by neural network and hierarchical attention network. Li [11] uses the multi-head self-attention to fully extract context representation. Xie [12] uses attention to obtain sentiment information contained in the sentiment words.

Function Words-Relevant Methods. In the early research based on rules or traditional machine learning approaches, adverbs and conjunctions were used to calculate the sentiment [3, 4, 13] or assist classifiers in predicting the sentiment polarity [5–8]. With the help of deep learning, Qian [1] uses the features of sentiment lexica, negative words, and degree adverbs and achieves good results with LSTM. Liang [14] uses conjunctions to segment phrases to construct graph structures to encode contextual information.

Although these studies have tried to utilize function words in sentiment analysis, their adverbs and conjunctions are mainly based on empirical results. Therefore, combining the existing sentiment lexical resources, this paper introduces the CFKB [15–19] into sentiment analysis. Thereinto, negative adverbs can change the polarity in the modification of sentiment words, degree adverbs can enhance or weaken the tendency, and conjunctions can change or even reverse the sentiment orientation of clauses to varying degrees.

2.2 Chinese Function Word Usage Knowledge Base

Function words are a primary means of expressing grammatical meaning in Chinese [20]. Unlike content words that can act independently as a sentence component, function words have no semantic meaning. They must be attached to content words or phrases to express grammatical meaning, tone, or sentiment. CFKB classifies function words into six categories: adverb, preposition, conjunction, auxiliary word, modality, and locality. It comprehensively describes function words in terms of their attributes, such as POS, definitions, example sentences, and usage descriptions.

CFKB, providing reliable lexical resources for Chinese language processing and semantic understanding, is widely used in the syntactic analysis [21, 22], information extraction [23], sentiment analysis [2], and other natural language processing tasks. Li [2] introduced degree adverbs, negative adverbs, and conjunctions in CFKB into sentiment analysis. Experimental results of the rule-based method are better than traditional machine learning methods, showing that function words have essential effects on sentiment word recognition and analysis. However, it still needs to be improved with deep learning methods.

3 Methodology

Based on the context representation obtained by ERNIE, FuncSA uses the sentiment knowledge obtained from lexical resources and CFKB to extract the sentiment-aware semantic information for sentiment analysis. The overall structure is shown in Fig. 1.

3.1 ERNIE Encoder

Pre-trained language models such as BERT [24] have achieved substantial progress in sentiment analysis. Like BERT, ERNIE [25] consists of multiple Transformer encoder layers. In addition, ERNIE is consistent with BERT in training by adding the mask [CLS] and [SEP] before and after the input text. [CLS] is the first token to capture the information representation of the context, which will be used for downstream tasks. [SEP] is at the end of the sentence, indicating the end.

However, ERNIE introduced phrases and entity knowledge in the pre-training stage. Specifically, ERNIE’s mask strategy mainly has three ways: First, it adopts

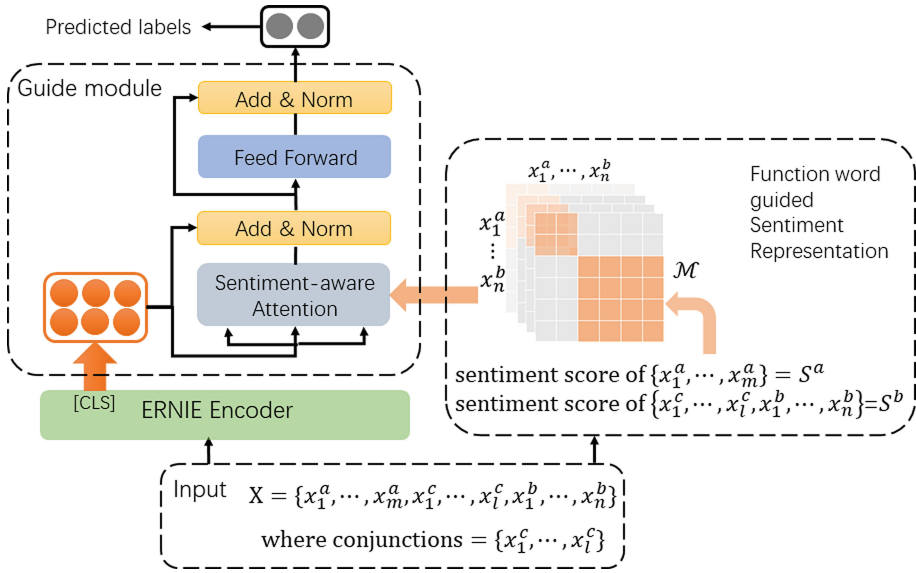


Fig. 1. Overview of FuncSA

the same way as BERT; that is, it masks and learns the expression of each piece in the input. Secondly, it randomly masks phrases to make use of word information. Finally, the entities in the input sentences are masked and predicted in the pre-training stage. ERNIE can learn the entities, phrases, and other external knowledge in the mass corpus to obtain a more reliable language representation through such a mask strategy. Therefore, we utilize ERNIE as the context encoder in Chinese sentiment analysis.

3.2 Function Words-Guided Sentiment Representation

Function words-guided Sentiment Representation is obtained through three steps. 1) Split the input sentence into clauses by conjunction. 2) Calculate the base sentiment score. 3) Weight the sentiment score by conjunctions.

Separate the Input Sentence into Clauses by Conjunctions. Conjunctions have complex and diverse functions and usages and can express various logical relations by connecting phrases, sentences, and texts. This paper selects four kinds of conjunctions in CFKB that express transition, progression, selection, and coordinate. The clauses connected by transition conjunctions usually have opposite sentiment orientations, and the clause after the conjunctions mainly determines the overall sentiment. Selective, progressive, and coordinating conjunctions tend to connect elements with the same polarity of sentiment, while progressive conjunctions tend to be followed by clauses with a more intense sentiment. The input sentence $X = \{x_1^a, \dots, x_m^a, x_1^c, \dots, x_\ell^c, x_1^b, \dots, x_n^b\}$ is divided into

Table 1. Examples of degree adverbs and the weights

Degree level	weight	Samples	number of words
extreme/most	4	beijia/倍加、feichang/非常、jiduan/极端	69
very	3	bushao/不少、fenwai/分外、gewai/格外	42
more	2	geng/更、jiao/较、yu/愈	37
-ish	0.5	lue/略、shao/稍、weimian/未免	29
insufficiently	0.1	buzenme/不怎么、qingdu/轻度、ruo/弱	12
over	5	chaoe/超额、guodu/过度、guofen/过分	30

two clauses $X_a = \{x_1^a, \dots, x_m^a\}$ and $X_{c+b} = \{x_1^c, \dots, x_\ell^c, x_1^b, \dots, x_n^b\}$ by conjunctions $X_c = \{x_1^c, \dots, x_\ell^c\}$ with the length of m and $\ell + n$, respectively.

Calculate the Base Sentiment Score. Then, we combine the HowNet, and the National Taiwan University Sentiment Dictionary(NTUSD)¹, and then remove the stop words in the collection and form FuncSA’s dictionary for calculating sentiment scores. In addition to sentiment words, the context of sentiment words, especially the degree and negative adverbs, is also essential to sentiment classification.

Degree adverbs significantly influence the strength of the sentiment orientation in sentences. The degree adverbs used in this paper come from HowNet. There are 219 °C words in Chinese, which can be divided into six categories according to degree level: “extreme/most, very, more, -ish, insufficiently, and over”. Each category is given a weight to calculate the sentiment score in sentences, and the weight is selected by experience. Examples of degree adverbs and the weights assigned to each category are shown in Table 1.

As for the negative adverbs, we screened all 50 negative adverbs in CFKB. In Chinese, there are double negation and multiple negation cases. When negation occurs even times, the original text logically has an affirmative meaning, and when negation occurs at odd times, the original text indicates a negative meaning. The number of negative adverbs in front of sentiment words is counted, and the sentiment value modified by odd negative adverbs is negative, while the value modified by even negative adverbs remains unchanged.

Based on the above steps, the sentiment scores of clauses X_a and X_{c+b} under the influence of degree adverbs and negative adverbs is S_a and S_{c+b} respectively.

Weight the Sentiment Score by Conjunctions. According to the conjunctions X_c , assign different weights w_1 and w_2 to S_a and S_{c+b} respectively, then the weighted sentiment scores of clauses X_a and X_{c+b} can be expressed as $S_a \times w_1$, $S_b \times w_2$. The weighted scores are assigned to each character in the clauses, and extended to the matrix $A_1 \in \mathbb{R}^{m \times m}$, $A_2 \in \mathbb{R}^{(\ell+n) \times (\ell+n)}$:

¹ <http://academiasinicanplab.github.io/>.

$$\begin{aligned}
A_1 &= w_1 \times S_a \times \begin{bmatrix} 1 \cdots 1 \\ \vdots \quad \vdots \\ 1 \cdots 1 \end{bmatrix} = \begin{bmatrix} w_1 S_a \cdots w_1 S_a \\ \vdots \quad \quad \quad \vdots \\ w_1 S_a \cdots w_1 S_a \end{bmatrix} \\
A_2 &= w_2 \times S_b \times \begin{bmatrix} 1 \cdots 1 \\ \vdots \quad \vdots \\ 1 \cdots 1 \end{bmatrix} = \begin{bmatrix} w_2 S_b \cdots w_2 S_b \\ \vdots \quad \quad \quad \vdots \\ w_2 S_b \cdots w_2 S_b \end{bmatrix}
\end{aligned} \tag{1}$$

Thus, the sentiment-aware representation M guided by function words can be expressed as:

$$M = \begin{bmatrix} A_1 & O \\ O & A_2 \end{bmatrix} \in \mathbb{R}^{(m+\ell+n) \times (m+\ell+n)} \tag{2}$$

3.3 Guide Module

FuncSA added one more Transformer encoder layer above ERNIE to integrate function words-guided sentiment representation. A Transformer encoder layer mainly comprises multi-head attention, feed-forward network, and Layer Norm. Self-attention learns the semantic representation of context by calculating the interaction between words. Multi-head self-attention expands feature space by calculating self-attention in different subspaces and improves implementation.

We integrate the sentiment-aware representation M into multi-head self-attention, which constitutes the sentiment-aware attention, to capture the interaction of the sentiment words and the context information. The guide module can learn implicit information in combination with sentiment knowledge in context representation obtained from the pre-trained model.

Specifically, for the context representation H , we apply function words-guided sentiment representation M on the QK^T to obtain self-attention:

$$Attention(Q, K, V) = Softmax\left(\frac{QK^T}{\sqrt{d_K}} + M\right)V \tag{3}$$

$$where \quad Q = HW^Q, K = HW^K, V = HW^V$$

where QK^T calculates the internal correlation between words in the clause. M is added to the attention score normalized by $\sqrt{d_K}$ and then the weighted attention score is obtained by multiplying $Softmax$ normalization by the V . Therefore, it can exert an influence on self-attention through sentiment information. Then, the attention scores from multiple subspaces are concatenation according to Eq. 4.

$$\begin{aligned}
MultiHead(Q, K, V) &= Concat(head_1, \dots, head_h)W^O \\
where \quad head_i &= Attention(Q_i, K_i, V_i)
\end{aligned} \tag{4}$$

According to Eq. 2, M is the diagonal block matrix so that attention will be limited within each clause. Each character only pays attention to the other in

the same clause because the sentiment or polarity has changed with the influence of conjunctions.

Furthermore, layer normalization and a feedforward network are used to accelerate the convergence and enhance the analyses and prediction of the model.

4 Experimental Settings

4.1 Datasets

In this paper, ChnSentiCorp [26], COAE2013², and NLPCC2014³ is selected to verify the performance of FuncSA.

ChnSentiCorp is an online shopping review dataset containing hotels, laptops, and books. To conduct a fair experimental comparison, this article follows the division of datasets in previous studies [25].

COAE2013 is from The Fifth Chinese Opinion Analysis Evaluation. There are 1004 positive reviews among the annotated data and 834 negative reviews. The dataset is divided into train set and test set according to the ratio of 9:1.

NLPCC2014 is from the sentiment classification with deep learning technology task on the 3rd CCF Conference on Natural Language Processing & Chinese Computing, utilizing data from Chinese product review websites, including books, DVDs, and electronic products reviews.

4.2 Baselines

We compare our model with the following baseline methods on both datasets.

- 1) RNNs or CNNs baselines: BiLSTM [27], BiLSTM+Att [27], TextCNN [28], DPCNN [29].
- 2) Vallina pre-trained models: BERT, BERT-WWM [30], RoBERTa [31], ERNIE.
- 3) ERNIE-based models: ERNIE+BiLSTM, ERNIE+BiGRU, and ERNIE+Att.

For the RNNs or CNNs baselines, we used the word vector pre-trained by the Sogou News corpus, running a maximum of 100 epochs with a batch size of 128. Adam optimizer was adopted with a learning rate of 1e-4. For BiLSTM and BiLSTM+Att, the dimension of the hidden layer was set to 128. For TextCNN and DPCNN based on CNN, the number of convolution kernels was 256.

For the pre-trained model baselines, we followed default settings, i.e., 12 layers of multi-head attention with the dimension of the hidden layer set to be 768. The batch size was set as 16, and the learning rate was 5E-5. Adam was used to optimize the cross-entropy loss function.

² <http://www.cipsc.org.cn/hytx/13.html#23>.

³ <http://tcci.ccf.org.cn/conference/2014/>.

Table 2. Performance comparison of models on four datasets.

Model	ChnSentiCorp-Dev		ChnSentiCorp-Test		COAE2013		NLPCC2014	
	Acc.(%)	F1.(%)	Acc.(%)	F1.(%)	Acc.(%)	F1.(%)	Acc.(%)	F1.(%)
BiLSTM	71.83	71.69	72.25	72.23	85.74	85.39	60.48	60.48
BiLSTM+Att	77.17	77.14	77.50	77.49	86.91	86.76	69.60	69.56
TextCNN	82.75	82.70	82.67	82.67	89.65	89.46	69.04	68.85
DPCNN	83.92	83.92	85.83	85.82	87.30	87.07	62.48	58.88
BERT	93.92	93.91	95.19	95.19	93.57	93.53	79.61	79.61
BERT-WWM	94.17	94.17	94.76	94.76	94.85	94.83	80.21	80.20
RoBERTa	94.00	94.00	94.76	94.76	95.04	95.04	79.57	79.56
ERNIE	94.26	94.25	95.27	95.27	95.77	95.74	80.89	80.88
ERNIE+BiLSTM	93.83	93.83	95.02	95.02	95.59	95.36	80.37	80.36
ERNIE+BiGRU	94.51	94.51	95.35	95.10	96.32	96.31	80.05	80.02
ERNIE+Att	94.34	94.34	95.44	95.44	96.69	96.67	80.69	80.67
FuncSA	94.43	94.42	96.20	96.20	97.43	97.41	81.37	81.35

5 Experimental Results

5.1 Main Results

The results are shown in Table 2, from which several observations can be obtained. First, the results of vanilla pre-trained models on all datasets are better than the optimal results of the RNNs or CNNs baselines. The pre-trained model can capture long-distance features better than RNNs and CNNs. Second, comparing these vanilla pre-trained models, we found that ERNIE reaches the best accuracy and F1 on both datasets. ERNIE could better adapt to the Chinese language and learn a more appropriate global representation.

Last but not least, comparing the ERNIE-based models, ERNIE+Att has an improvement compared with ERNIE+BiLSTM and ERNIE+BiGRU because attention can adjust the global representation learned by ERNIE to focus on critical areas in the context. FuncSA added sentiment knowledge guided by function words and integrated sentiment knowledge with context information through one more Transformer encoder layer. Compared with ERNIE+Att, FuncSA achieved better results in accuracy and F1, which fully demonstrates the effectiveness of this model.

5.2 Ablation Study

In order to explore the directive role of different classes of function words and lexical information on sentiment-aware attention, we discuss the performance of FuncSA with or without function words-guided sentiment representation M and different function words combination. The results are shown in Table 3.

Results show that although accuracy and F1 of FuncSA are decreased by 0.16% and 0.17% compared with w/o M on the ChnSenticorp-Dev, which may

Table 3. Ablation study.

Model	Lexicon	Adv.	Conj.	ChnSentiCorp-Dev		ChnSentiCorp-Test		COAE2013		NLPCC2014	
				Acc.(%)	F1.(%)	Acc.(%)	F1.(%)	Acc.(%)	F1.(%)	Acc.(%)	F1.(%)
FuncSA	✓	✓	✓	94.43	94.42	96.20	96.20	97.43	97.41	81.37	81.35
w/o <i>M</i>	×	×	×	94.59	94.59	94.59	94.35	96.88	96.86	81.05	81.04
w/o conj. & adv.	✓	×	×	94.26	94.26	95.52	95.52	97.06	97.05	80.33	80.31
w/o conj.	✓	×	✓	94.09	94.08	95.86	95.86	96.32	96.31	80.73	80.72
w/o adv.	✓	✓	×	94.34	94.34	95.78	95.78	97.06	97.04	80.81	80.80

result from unexpected noise caused by the introduction of the sentiment-aware representation. However, FuncSA is better than others in most datasets, illustrating the function words-guided knowledge can effectively guide the attention to focus on sentiment-specific areas.

Comparing w/o conj. & adv. with w/o conj., it is found that in the ChnSenticorp-Test, and NLPCC2014 datasets, accuracy, and F1 of w/o conj. are all higher than that of W/O conj. & adv., which indicates that adverbs are indispensable.

Comparing w/o conj. & adv. with w/o adv., which both considered sentiment words but did not consider adverbs, the results of w/o adv. were better than the accuracy of w/o conj. & adv. in all datasets, indicating that there is indeed an association between conjunctions and sentiment words in Chinese text sentiment analysis.

Furthermore, by comparing w/o *M* with the other three variants, it was found that the results of other variants were better than the results of w/o *M* in the ChnSenticorp-Test, COAE2013, and NLPCC2014. FuncSA makes full use of the sentiment information of lexicons, and the modification of conjunctions and adverbs, thus improving the accuracy of classification.

5.3 Visualization

In order to intuitively compare the semantic expression ability of baseline ERNIE with FuncSA in Chinese text, we use BertViz [32], an open-source attention visualization tool for pre-trained models. For example, the text “房间隔音和餐饮服务都不好(The sound insulation of the room and catering service are not good.)” is the case for the attention view, as shown in Fig. 2. Since different layers have different attention modes, this paper selects the view of all attention subspaces at the last layer, and the color blocks at the top of Fig. 2 distinguish the attention representation of each 12 subspaces. The intensity of the attention distribution in each subspace determines the color saturation of the attention view on the right side of each row.

The leftmost column of the attention head view contains the model input with two identifying masks [CLS] and [SEP] and the comment text. The [CLS] represents the hidden state output of the input, so the attention distribution from the [CLS] to other positions in the sequence was tracked. It can be seen that FuncSA’s multiple subspaces pay more attention to the text containing

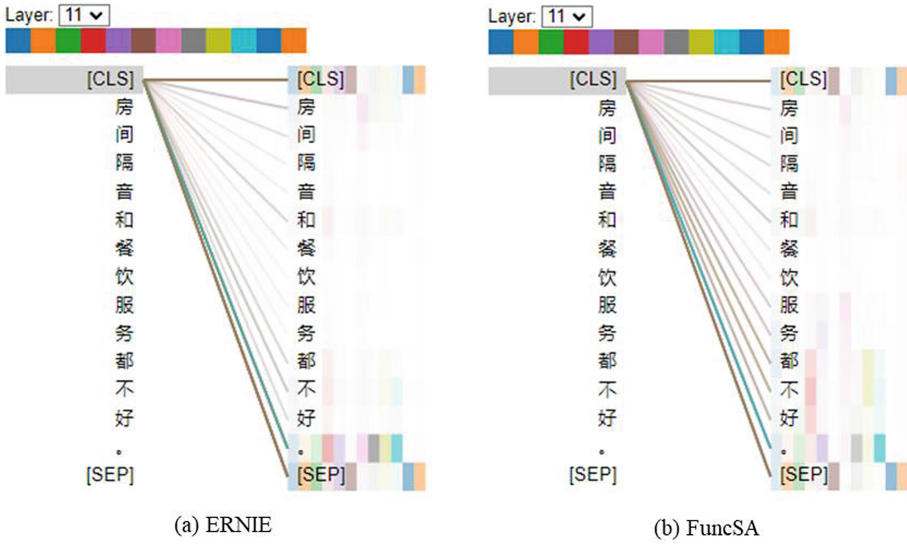


Fig. 2. Attention detail view of ERNIE and FuncSA: (a) ERNIE; (b) FuncSA

negation tendencies “dou buhao/都不好(are not good)”, and it is easier to make correct sentiment polarity classification than ERNIE.

6 Conclusion

This paper mainly introduces FuncSA, a sentiment analysis model which integrates function words-guided sentiment-aware attention. FuncSA integrates discrete sentiment lexical information with degree adverbs, negative adverbs, and conjunctions in CFKB. Sentiment-aware attention is then introduced to assist the pre-trained model in sentiment classification. Experimental results show that the proposed FuncSA can improve the results of Chinese sentiment analysis. This study reveals the validity of sentiment lexical information influenced by function words and verifies the role of adverbs and conjunctions in guiding sentiment changes.

References

1. Qian, Q., Huang, M., Lei, J., Zhu, X.: Linguistically regularized LSTM for sentiment classification. In: ACL 2017–55th Annual Meeting of the Association for Computational Linguistics, Proceedings of the Conference (Long Papers) (2017). <https://doi.org/10.18653/v1/P17-1154>
2. Li, Y.: Microblog Emotional Dictionary Built and Application on Sentiment Analysis of Microblog. Zhenzhou University (2014)

3. Shi, W., Fu, Y.: Microblog short text mining considering context: a method of sentiment analysis. *Jisuanji Kexue/Comput. Sci.* **48**(6A), 158–164 (2021). <https://doi.org/10.11896/jsjcx.210200089>
4. Lipenkova, J.: A system for fine-grained aspect-based sentiment analysis of Chinese. In: *ACL-IJCNLP 2015–53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing, Proceedings of System Demonstrations* (2015). <https://doi.org/10.3115/v1/p15-4010>
5. Ku, L.W., Huang, T.H., Chen, H.H.: Using morphological and syntactic structures for Chinese opinion analysis. In: *EMNLP 2009 - Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing: A Meeting of SIGDAT, a Special Interest Group of ACL, Held in Conjunction with ACL-IJCNLP 2009* (2009). <https://doi.org/10.3115/1699648.1699672>
6. Li, C., Wu, H., Jin, Q.: Emotion classification of Chinese microblog text via fusion of BoW and eVector feature representations. In: *Communications in Computer and Information Science* (2014)
7. He, Y., Zhao, S., He, L.: Micro-text emotional tendentious classification based on combination of emotion knowledge and machine-learning algorithm. *J. Intell.* **37**(5), 189–194 (2018)
8. Xu, J., Ding, Y., Wang, X.: Sentiment classification for Chinese news using machine learning methods. *J. Chin. Inf. Process.* **21**(6), 95–100 (2007)
9. Ambartsoumian, A., Popowich, F.: Self-attention: a better building block for sentiment analysis neural network classifiers. In: *WASSA 2018–9th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis, Proceedings of the Workshop* (2018). <https://doi.org/10.18653/v1/P17>
10. Cheng, Y., Ye, Z., Wang, M., Zhang, Q., Zhang, G.: Chinese text sentiment orientation analysis based on convolution neural network and hierarchical attention network. *J. Chin. Inf. Process.* **33**(1), 133–142 (2019)
11. Li, Z., Chen, L., Zhang, S.: Chinese text sentiment analysis based on ELMO and Bi-SAN. *Appl. Res. Comput.* **38**(8), 2303–2307 (2021)
12. Xie, R., Li, Y.: Text sentiment classification model based on BERT and dual channel attention. *Shuju Caiji Yu Chuli/J. Data Acquis. Process.* **35**(4), 642–652 (2020). <https://doi.org/10.16337/j.1004-9037.2020.04.005>
13. Liu, Y., Ju, S., Wu, S., Su, C.: Sentiment classification of Chinese texts based on emotion dictionary and conjunction. *J. Sichuan Univ. (Nat. Sci. Edn.)* **52**(1), 57–62 (2015)
14. Liang, S., Wei, W., Mao, X., Wang, F., He, Z.: BiSyn-GAT+: Bi-Syntax Aware Graph Attention Network for Aspect-based Sentiment Analysis. (2022)
15. Zan, H., Zhang, K., Zhu, X., Yu, S.: Research on the Chinese function word usage knowledge base. *Int. J. Asian Lang. Process.* **21**(4), 185–198 (2011)
16. Zan, H., Zhang, K., Chai, Y., Yu, S.: Studies on the functional word knowledge base of modern Chinese. *J. Chin. Inf. Process.* **21**(5), 107–111 (2007)
17. Zhan, H., Zhu, X.: Mianxiang Ziran Yuyan Chuli de Hanyu Xuci Yanjiu yu Guangyi Xuci Zhishiku Goujian [Research on Chinese Function words for Natural Language Processing and Construction of generalized Function words Knowledge Base]. *Contemp. Linguist.* **11**(2), 124–135 (2009)
18. Zhang, K., Zan, H., Chai, Y., Han, Y., Zhao, D.: Construction and application of the Chinese function word usage knowledge base. *Int. J. Knowl. Lang. Process.* **4**, 32–42 (2013)
19. Zhang, K., Zan, H., Chai, Y., Han, Y., Zhao, D.: Survey of the Chinese function word usage knowledge base. *J. Chin. Inf. Process.* **29**(3), 1–8 (2015)

20. Huang, B., Liao, X.: *Xiandai Hanyu [Modern Chinese]*. Higher Education Press, Beijing (2011)
21. Zan, H., Zhang, J., Lou, X.: Studies on the application of Chinese functional words usages in dependency parsing. *J. Chin. Inf. Process.* **27**(5), 35–43 (2013)
22. Mu, L., Pang, Y., Zan, H.: Studies on the usage of preposition ZAI in phrase structure syntactic parsing (2014)
23. Zan, H., Zhang, T., Lin, A.: Research on event information extraction based on preposition's usages. *Comput. Eng. Design* **34**(7), 2570–2574 (2013)
24. Devlin, J., Chang, M.W., Lee, K., Toutanova, K.: BERT: pre-training of deep bidirectional transformers for language understanding. In: *NAACL HLT 2019–2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies - Proceedings of the Conference* (2019)
25. Sun, Y., et al.: ERNIE: enhanced representation through knowledge integration (2019)
26. Tan, S., Zhang, J.: An empirical study of sentiment analysis for Chinese documents. *Exp. Syst. Appl.* **34** (2008). <https://doi.org/10.1016/j.eswa.2007.05.028>
27. Zhang, D., Wang, D.: *Relation Classification via Recurrent Neural Network* (2015)
28. Kim, Y.: Convolutional neural networks for sentence classification. In: *EMNLP 2014–2014 Conference on Empirical Methods in Natural Language Processing, Proceedings of the Conference* (2014). <https://doi.org/10.3115/v1/d14-1181>
29. Johnson, R., Zhang, T.: Deep pyramid convolutional neural networks for text categorization. In: *ACL 2017–55th Annual Meeting of the Association for Computational Linguistics, Proceedings of the Conference (Long Papers)* (2017). <https://doi.org/10.18653/v1/P17-1052>
30. Cui, Y., et al.: Pre-training with whole word masking for Chinese bert (2019)
31. Liu, Y., et al.: RoBERTa: a robustly optimized BERT pretraining approach (2019)
32. Vig, J.: A multiscale visualization of attention in the transformer model. In: *ACL 2019–57th Annual Meeting of the Association for Computational Linguistics, Proceedings of System Demonstrations* (2019). <https://doi.org/10.18653/v1/p19-3007>