

International Journal of Computational Intelligence Studies

ISSN online: 1755-4985 - ISSN print: 1755-4977
<https://www.inderscience.com/ijcistudies>

Aspect-level sentiment classification with emotional keywords attention network

Zhifeng Yuan, Jin Yuan

DOI: [10.1504/IJCISTUDIES.2026.10076986](https://doi.org/10.1504/IJCISTUDIES.2026.10076986)

Article History:

Received:	21 March 2023
Last revised:	20 October 2025
Accepted:	22 November 2025
Published online:	18 March 2026

Aspect-level sentiment classification with emotional keywords attention network

Zhifeng Yuan

Puyang Vocational and Technical College,
Economic and Technological Development Zone,
No. 249, West Huanghe Road,
Puyang City, Henan Province, China
Email: zfyuan@1970@gmail.com

Jin Yuan*

Lenovo,
Software Park, Houchangcun Road,
Haidian District, Beijing, China
Email: yuanjin@seu.edu.cn

*Corresponding author

Abstract: Aspect-level sentiment classification (ASC) aims to uncover the sentiment polarity expressed towards specific entities in market analysis. Despite its popularity, two fundamental obstacles persist: 1) manual annotation is notoriously labour-intensive, resulting in data-scarce regimes where models struggle to acquire sufficient knowledge; 2) accurate sentiment inference demands rich prior semantic knowledge that is difficult to acquire even when large corpora are available. To address these challenges, we first automatically identify emotion-bearing keywords for each aspect and summarise their distributional properties. We then introduce a convolutional neural network enhanced with an attention mechanism that explicitly highlights these keywords through an emotion-keyword mask, thereby guiding the model to focus on sentiment-relevant context. Extensive experiments on standard SemEval benchmarks show that the proposed mask consistently improves performance and that our overall approach achieves competitive results against recent strong baselines.

Keywords: attention network; sentiment classification; natural language processing.

Reference to this paper should be made as follows: Yuan, Z. and Yuan, J. (2026) 'Aspect-level sentiment classification with emotional keywords attention network', *Int. J. Computational Intelligence Studies*, Vol. 13, No. 5, pp.1–13.

Biographical notes: Zhifeng Yuan studied Management Engineering at China University of Petroleum from September 1988 to July 1992, and he also graduated with a Bachelor's degree in July 1992. From February 2007 to

June 2009, he studied Business Administration at Zhengzhou University and obtained a Master of Business Administration (MBA) in June 2009. His main research interests are user management and sentiment analysis.

Jin Yuan studied Computer Science at Hunan University from September 2013 to July 2017, and graduated with a Bachelor’s degree in July 2017. From September 2020 to December 2024, he studied Computer Science at Southeast University and obtained a PhD in Computer Science in December 2024. His main research interests are artificial intelligence and agent.

1 Introduction

The goal of aspect-level sentiment classification (ASC) is to determine the sentiment polarity expressed toward specific entities or their facets in a sentence. Previous studies (Pang et al., 2008; Liu, 2012) further divide ASC into two subtasks: aspect category sentiment analysis (ACSA) and aspect term sentiment analysis (ATSA). For the sentence “the pizza is the best”, ACSA assigns polarity to the predefined category *food*, whereas ATSA assigns polarity to the explicit aspect term *pizza*. Empirical results show that, when trained on datasets of comparable size with the same model, the accuracies of ACSA and ATSA are almost identical; consequently, this paper focuses on the ATSA subtask.

Aspect-level datasets suffer from three critical shortcomings that degrade model performance. First, textual noise – misspellings, abbreviations, and informal expressions – disrupts feature extraction. Second, hard cases containing opposing sentiment polarities within a single sentence challenge semantic understanding. Third, and most severe, is data scarcity: constructing fine-grained annotations demands extensive expertise and labour, making large-scale datasets prohibitively expensive.

Recent state-of-the-art approaches rely on recurrent architectures augmented with attention mechanisms. While these models attain competitive accuracy, deepening the network yields only marginal gains and incurs prohibitive latency, limiting their use in real-time sentiment services. Moreover, their performance degrades markedly on sentences that convey complex or conflicting emotions; representative easy and hard cases are illustrated in Table 1.

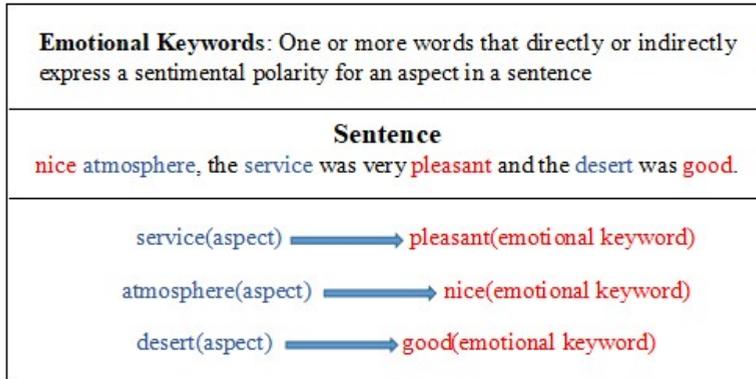
Table 1 Examples for easy cases and hard cases

<i>Type</i>	<i>Sentence</i>	<i>Aspects</i>
Easy	It rarely works and when it does it’s incredibly slow.	Work
Easy	The design and atmosphere is just as good.	Design, atmosphere
Hard	They cost more, but they more than make up for it in speed.	Cost
Hard	The menu is limited but almost all of the dishes are excellent.	Menu, dishes

To address these limitations, we first operationalise the concept of emotional keywords (Figure 1) and statistically validate their discriminative power on a large-scale corpus. Building on this insight, we propose a convolutional architecture enhanced with

attention: after feature embedding and the attention layer, an emotional-keyword mask explicitly guides the model toward sentiment-relevant signals. Extensive experiments across four strong baselines and multiple datasets demonstrate consistent gains, with particularly pronounced improvements on hard cases.

Figure 1 Definition of emotional keywords and an example showing what is emotional keywords in a sentence (see online version for colours)



2 Related work

A lot of work related to sentiment analysis has been proposed. In this section, we mainly discuss the work related to aspect-level sentiment classification according to different methods.

Traditional solutions usually focus on designing an emotion package or emotion dictionary, and then design some classifiers, such as SVM, to classify emotion information (Jiang et al., 2011; Mohammad et al., 2013; Kiritchenko et al., 2014). Nevertheless, these approaches are always labour-intensive, and their performance largely depends on emotional packages or emotional dictionaries. Besides, tree-based recursive neural networks were also used for sentiment classification (Socher et al., 2013; Tai et al., 2015; Wang et al., 2016a). Due to the poor anti-noise ability, these models are easy to parse errors, especially for noisy text, so their performance on the review dataset is limited.

In 1997, Hochreiter and Schmidhuber proposed the long short-term memory network (LSTM), and Mikolov et al. (2010) confirmed the effectiveness of the recurrent neural network-based method for language tasks. Subsequently, Bahdanau et al. (2014) proposed the attention mechanism in 2014, and since then, methods based on LSTM and attention mechanism have gradually become the mainstream model of linguistics tasks. Recently, some works (Vo and Zhang, 2015; Wang et al., 2016b; Tang et al., 2016; Ma et al., 2017; Chen et al., 2017) gradually proposed some solutions to the emotion classification problem of LSTM based on attention mechanism. Wang et al. (2016b) proposed aspect embedding and LSTM with attention mechanism to solve aspect category sentiment classification. Tang et al. (2016) adopted a memory network approach to explore fine-grained sentiment classification, which builds knowledge in some vectors to understand the entire context. Ma et al. (2017) proposed an interactive

LSTM model between target and context, which modelled the target separately and made it interact with the sentence model. Chen et al. (2017) proposed the RAM model, which designed a weighted LSTM model by emphasising the location information of the target, and adopted multiple attention layers.

Other work attempts to solve this problem by cross-domain (Blitzer et al., 2007; Li et al., 2009; Pan et al., 2010; Glorot et al., 2011; Bollegala et al., 2012; Zhang et al., 2017; He et al., 2018; Li et al., 2019). They tried to improve the accuracy of sentiment classification by aligning sentiment classification tasks with different granularities. He et al. (2018) tried to learn from text-level data and use it for aspect-level sentiment classification, which simultaneously trained document-level and aspect-level tasks by sharing their embedding layer and LSTM layer and using both losses to solve the problem of insufficient datasets. Li et al. (2019) tried to learn knowledge from the ACSA task to the ATSA task. They proposed a C2F attention network to reduce the difference between those two tasks and added a separate classification layer for semantic alignment.

Many works have also tried to use convolutional neural networks (CNN) to solve the problem of sentiment classification but mainly based on text-level or sentence-level datasets (Kalchbrenner et al., 2014; Kim, 2014; Conneau et al., 2016). Xue and Li (2018) proposed a convolutional neural network with a gating mechanism for aspect-level sentiment classification, which used gated tanh-ReLU units (GTRU) to help convolutional neural networks to get features and finally had a stable performance and shorter training time.

3 Emotional keywords attention network

3.1 Convolutional neural network with attention mechanism

Based on a convolutional neural network model (Xue and Li, 2018), which used a simple and efficient convolutional layer to finish the sentiment classification, we tried to build our model. The most significant limitation of this baseline model is the lack of accuracy and stability. We consider this is because the model does not have enough capacity to obtain information stably, which may be related to the randomness of the convolutional layer. Therefore, we consider solving the current situation. Next, we will detail our model.

First, the input is a sentence $s = \{s_1, s_2, \dots, s_n\}$ of length n , an aspect target $a = \{a_1, a_2, \dots, a_m\}$ of length m , and an emotional polarity p . We use the word segmentation tool to segment the word first, dividing the sentence into n dependent words, dividing the aspect target word into m independent words. Subsequently, we find the target word position from aps to ape in the sentence and use the word segmentation tool to save the part-of-speech information of each word. To save the data more effectively, we use the word dictionary established by the word segmentation tool to map the sentence s and target word a to s_m and a_m after the word segmentation. Finally, the mapped sentence, target word, part-of-speech of sentence, target position, and emotional polarity are sent to CNN-based model.

In our CNN model, the sentences and target words of the same batch are first aligned to eliminate the problem of different lengths that are difficult to train. The sentence s_m and the target word a_m then go through an embedding layer to generate embedding

vector s_v and a_v . The embedding vector has a dimension of (n, D) , where D is the defined embedding length. The embedding layer usually loads some pre-trained files, and words that are not in the file are randomly initialised into a vector of the same length. The embedding sentence s_v will continue to pass through an attention layer and get s_{va} . We use multi-head attention (Vaswani et al., 2017) to further complete the feature extraction and model stabilisation. Then, we obtain the mask by the position and part-of-speech information and let the feature s_{va} multiple this mask, which could emphasise the key information and get s_{vam} . Then, s_{vam} will concatenate with s_v and get the final embedded feature vector f_v .

For embedded feature vector f_v , we use one-dimensional convolutional layer with an activation function to continue extracting features:

$$f_x = \tanh(\text{conv}(f_v)) \quad (1)$$

Similarly, the target word vector a_v also go through the convolutional layer, the pooling layer, and the fully connected layer and then add to the convoluted f_v . The result passes through the activation function:

$$a_v = \text{relu}(\text{conv}(a_v)) \quad (2)$$

$$a_v = \text{maxpool}(a_v) \quad (3)$$

$$a_v = \text{fc}(a_v) \quad (4)$$

$$f_y = \text{relu}(a_v + \text{conv}(f_v)) \quad (5)$$

Then we multiply f_x and f_y and pass the pooling and fully connected layers to get the result r :

$$f = \text{maxpool}(f_x * f_y) \quad (6)$$

$$r = \text{fc}(f) \quad (7)$$

Finally, we use softmax for classification:

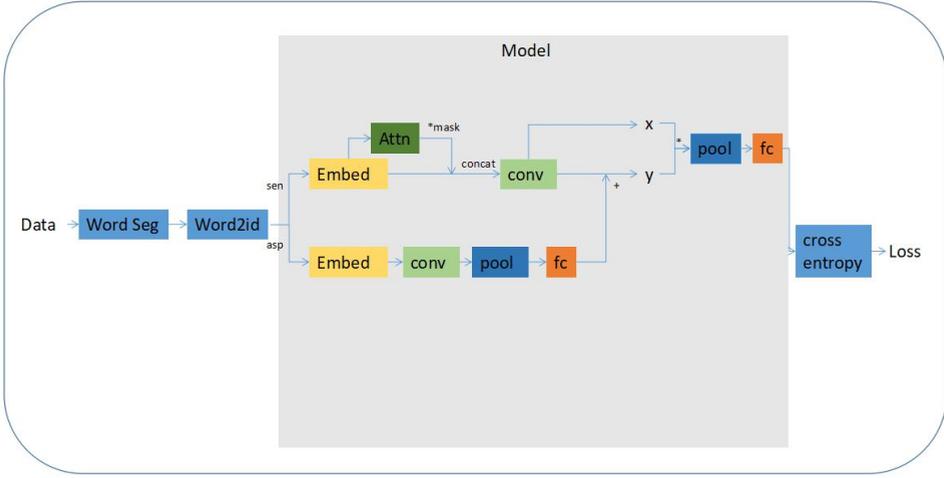
$$p = \text{softmax}(r) \quad (8)$$

And calculate the cross entropy as a loss function:

$$L = - \sum_{i \in D} (p_i(c_i)) \quad (9)$$

where c_i denotes the true label for sample i , and $p_i(c_i)$ denotes the probability of true label, and D denotes the whole dataset.

Our full model is shown in Figure 2.

Figure 2 Our full model architecture (see online version for colours)

3.2 Emotional keyword mask

After establishing a basic model structure, we found that the performance improvement of sentiment was limited. Then, we observed the review dataset we used and found that for each piece of data, there is a corresponding emotional keyword that can directly or indirectly point to emotional tendencies. Models often cannot reasonably extract these keywords, especially when there are multiple words with different emotional polarity in some sentences. Therefore, we decided to use an emotional keyword mask to assist the judgment of the model.

We first looked at some data and found that sentimental keywords tend to be closer to the terms, so we first defined the rules for distance. We define the start position of the aspect word in the sentence as s , the end position as e , the length of the entire sentence is l , and the mask value of the i^{th} bit is $mask_i$.

$$mask_i^* = \begin{cases} 1.0 - \frac{s-i}{l}, & \text{if } i < s \\ 1.0, & \text{if } i \geq s \text{ and } i < e \\ 1.0 - \frac{i-e}{l}, & \text{if } i \geq e \end{cases} \quad (10)$$

In addition, we also find that most of the emotional keywords are adjectives, verbs, and adverbs. To this end, we try to assign different weights to words with different part-of-speech and define the part-of-speech of the i^{th} word p_i .

$$mask_i^* = \begin{cases} 1.0, & \text{if } p_i \text{ in [adjectives, verbs, adverbs]} \\ 0.01, & \text{if } p_i \text{ not in [adjectives, verbs, adverbs]} \end{cases} \quad (11)$$

With a simple experiment, we found that this mask could improve model performance. However, the previous weights are only set through observation, and we cannot consider it as the optimal value of this mask. Therefore, we decided to select a part from the dataset for labelling and consider the characteristics of emotional keywords in part-of-speech and location from a statistical perspective.

We randomly selected thousands of data from the dataset to annotate. We used the word segmentation tool to segment the data first, and then for each data (including an aspect word and a sentence), we labelled the corresponding emotional keywords position. Subsequently, we first calculated the distance between the emotional keywords and the aspect words. The number of times that the emotional keyword with a distance i in the entire labelled dataset is defined as C_i , and the number of all keywords is N . Then the mask weight with distance i is:

$$mask_i^* = \frac{C_i}{N} \quad (12)$$

We also counted the number of occurrences of various part-of-speech in all emotional keywords. For part-of-speech p_i , we defined it as emo_{p_i} . However, since different part-of-speech have different proportions in sentences, it is not accurate to define the mask only by the number of occurrences of different part-of-speech. Therefore, we also counted the number of occurrences of various part-of-speech in the entire labelled dataset. For the part-of-speech p_i , the overall number of occurrence is defined as all_{p_i} and the mask value should be:

$$mask_i^* = \frac{emo_{p_i}}{all_{p_i}} \quad (13)$$

4 Experiments

4.1 Dataset and experimental setting

We tested our models on the SemEval dataset, including SemEval2014, SemEval2015, and SemEval 2016. The SemEval dataset mainly includes two categories, namely restaurant, and laptop. Table 2 shows more information about SemEval14. Each piece of data contains an aspect word, a sentence, and an emotional polarity. Emotional polarity includes four categories: positive, negative, neutral, and conflict. Unlike some previous works to remove the emotional label *conflict*, we used all the raw data for the test to evaluate the model’s performance in dealing with the four classification problems. *Conflict* means that people have two or more different emotional polarities towards an entity, such as the sentence “dishes there are very delicious, but the price is a little high”. People have conflicting emotional polarities towards *dishes*. Besides, we also introduced hard test sets. The hard test set is a set of sentence containing two different emotional polarities, which can largely reflect the model’s ability to understand the complex sentences.

Table 2 Details in SemEval14 dataset

Type	Trainset	Testset	Testset (hard)
Restaurant	3,693	1,134	245
Laptop	2,358	654	108

For embedding vectors, we used GloVe (Pennington et al., 2014) pre-trained file, which was trained on 840 billion tokens, to initialised all the vectors including target

and sentences with 300-dimensions and used the spaCy (Honnibal et al., 2020) word segmentation tool for word segmentation and POS tag. We also used 32 batch size and model were trained by AdaGrad optimiser by minimising cross-entropy loss function with 0.01 learning rate.

4.2 Effectiveness of emotional keywords

As mentioned in Section 3.2, in order to further explore the features of emotional keywords, we selected some data from SemEval2014. We randomly selected 3,693 entries from the restaurant dataset and 2,358 entries from the laptop dataset. We first use spaCy word segmentation tool to segment each sentence and then labels the serial number corresponding to the emotional keyword.

Table 3 Top 10 distance between emotional keywords and aspect words in restaurant and laptop dataset

<i>Lap dist</i>	<i>Lap prob</i>	<i>Rest dist</i>	<i>Rest prob</i>
2	12.4%	3	10.4%
3	10.5%	2	9.4%
1	9.8%	1	8.5%
4	8.2%	4	8.0%
5	6.9%	5	7.7%
6	6.6%	6	6.3%
7	5.4%	8	6.1%
8	4.6%	7	6.0%
11	4.4%	11	5.6%
9	3.7%	9	3.7%

Table 4 Top 7 part-of-speech ratio of emotional keywords in restaurant and laptop dataset

<i>Lap pos</i>	<i>Lap prob</i>	<i>Rest pos</i>	<i>Rest prob</i>
PART	18.2%	PART	12.8%
VERB	9.9%	ADJ	9.1%
ADJ	9.0%	VERB	8.3%
SCONJ	7.6%	AUX	6.1%
ADV	7.5%	ADV	5.5%
ADP	7.1%	NOUN	5.0%
AUX	6.7%	ADP	4.9%

Table 5 Emotional keywords mask comparisons

<i>Model</i>	<i>Accuracy</i>
CNN w/o attn w/o mask	64.11
CNN w/ attn w/o mask	65.55
CNN w/o attn w/ mask	87.41
CNN w/ attn w/ mask	88.13

After completing the labelling, we first counted the distance information of all the labelled emotional keywords and aspect words and recorded the distances where the distance appeared more than 2%. The specific results are shown in Table 3.

This result proves the correctness of our speculation. Words closer to aspect words in the distance are more likely to become emotional keywords. Besides, we can also find that this rule could apply to different datasets. Comparing the laptop and restaurant datasets, we can find that the probability of emotional keywords appearing at the same distance is extremely close. Finally, we normalised the position weights and added an emotional keyword mask.

We then explored the number of times that different part-of-speech appear in emotional keywords, and count the number of times the corresponding part-of-speech appears in the entire sentence to calculate the ratio, as shown in Table 4.

Among them, PART means negative word, ADJ means adjective, VERB means verb, ADV means adverb, ADP means preposition, AUX means auxiliary verb, and CONJ means conjunction. By observation, we can find that this is consistent with our previous speculative verbs, adjectives, and adverbs (including negative words). Then, we normalise the value of the part-of-speech and add it to the emotional keyword mask.

To further illustrate the importance of emotional keywords, we have divided the labelled 3,693 restaurant data into 80% of the training set and 20% of the test set, and directly read our labelled keyword mask to test. The detailed results are shown in Table 5. This table shows that if we can know the position of the emotional keywords correctly, the accuracy of sentiment analysis will be largely improved. From this, we can infer that adding the specific information of emotional keywords to the model is also helpful to the model. In addition, we can find that the version with the attention mechanism has higher results than without attention mechanism, which also partially illustrates the effectiveness of the attention mechanism for the convolutional neural network.

4.3 Model comparison

Table 6 shows the model comparison of our model with attention mechanism, our model without attention mechanism, and four benchmarks:

- *ATAE* (Wang et al., 2016b): it uses the LSTM model with aspect embedding called AE-LSTM, and adds the attention layer to get ATAЕ-LSTM
- *IAN* (Ma et al., 2017): it uses two LSTM to deal with aspects and sentences respectively, and makes them interact to acquire each other’s knowledge
- *GCAE* (Xue and Li, 2018): it uses a convolutional neural network with a gated mechanism to solve the sentiment classification problem
- *RAM* (Chen et al., 2017): it uses multiple layers of attention and uses a weighted memory mechanism to capture long-term features.

These models are trained and tested on the SemEval14 dataset, which includes restaurant and laptop dataset. The test set contains both a standard test set and a hard test set. Since the models were oscillating, we trained and tested each model five times and recorded the accuracy intervals of different models. In order to compare the differences of models

more accurately and reduce the differences caused by hyperparameters, all models were trained with the same setting in Section 4.1.

Table 6 Comparison of accuracy between models on different datasets

<i>Model</i>	<i>Laptop</i>	<i>Laptop hard</i>	<i>Restaurant</i>	<i>Restaurant hard</i>
ATAE-LSTM	58.18 ± 0.38	40.28 ± 1.39	67.24 ± 0.84	50.41 ± 2.25
IAN	66.88 ± 1.25	41.67 ± 3.13	73.84 ± 1.79	50.90 ± 2.24
GCAE	66.13 ± 1.91	45.84 ± 4.17	66.59 ± 1.15	46.76 ± 2.32
RAM	68.65 ± 0.62	48.15 ± 1.85	76.37 ± 0.44	56.53 ± 1.02
OUR MODEL WITH ATTN	67.89 ± 1.22	56.85 ± 3.15	76.15 ± 0.31	60.82 ± 1.23
OUR MODEL WITHOUT ATTN	68.58 ± 0.84	55.09 ± 1.39	76.41 ± 0.57	60.21 ± 1.84

First, compared with our baseline GCAE, the proposed model delivers markedly more stable performance, indicating that it can partially mitigate the volatility caused by the randomness of convolutional neural networks. Secondly, it achieves the highest accuracy on both the restaurant and laptop benchmark datasets, demonstrating strong and consistent capability in understanding relatively simple sentences. Although the RAM model also performs excellently on these benchmarks, our architecture is substantially simpler. Specifically, our model outperforms the previous best-reported results by about 8.7% on the restaurant dataset and by about 4.3% on the laptop dataset. On the hard test set, it surpasses the GCAE baseline by more than 10%, evidencing its superior accuracy. Furthermore, the attention-enhanced variant consistently outperforms its non-attention counterpart on the hard test set, underscoring the effectiveness of the attention mechanism when handling complex sentences.

4.4 Ablation studies

In order to better verify the performance of each module, we tested the original CNN, the network only with the attention mechanism, the network with only the mask, and the final model. The results are shown in Table 7.

Table 7 Ablation studies for our models

<i>Model</i>	<i>Lap</i>	<i>Lap hard</i>	<i>Rest</i>	<i>Rest hard</i>
CNN	66.15	45.19	74.81	53.06
Only Attn	66.79	46.85	74.62	55.02
Only mask	68.47	55.19	76.33	60.41
Entire model	68.38	55.74	76.19	60.98

We can compare and find that the mask mechanism significantly improves the accuracy of the standard test set and hard test set. The attention mechanism has a definite increase in the hard test set. But in the standard test set, the attention model sometimes slightly lower than the version without attention. We think that attention is helpful for understanding complex contexts, but for some simple sentences, attention may also overfitting.

In addition, the dataset itself may also affect the performance of the model. Therefore, we transferred our method to a new dataset composed of SemEval15 and SemEval16 for testing. Since this dataset is larger than SemEval14, we can also test the performance of the model when dealing with large amounts of data. The specific results are shown in Table 8.

Table 8 Ablation studies in SemEval15 and SemEval16

<i>Model</i>	<i>Rest new</i>	<i>Rest new hard</i>
CNN	88.97	71.89
Only Attn	88.85	72.69
Only mask	89.08	75.3
Entire model	88.85	76.71

From Table 8, we can find that the results of the four models are very close to the standard test set. Only the mask version is slightly higher. This shows that when the amount of data is sufficient, the convolutional neural network has strong adaptability, and it can well understand some simple samples. For the hard test set, we can find that both the version with attention mechanism and with mask mechanism can bring some improvement. When we use both the attention and the mask mechanism, the result of the hard test set can be about 5% higher than that of the naive convolutional network. This shows that our network is adaptable and that it can achieve excellent performance in extended datasets.

5 Conclusions and future work

In this paper, we propose a convolutional neural network model with an attention mechanism that uses an emotional keyword mask to help the model obtain critical information about emotions, and finally confirms the validity of the model. In the future, we plan to train an additional network to predict sentimental keywords. If we can accurately predict sentimental keywords, the accuracy of the task of sentiment classification will be significantly improved. In fact, in this work, we have tried to train such a network, but we finally gave up due to the insufficient labelled data and the low accuracy of the model. Therefore, in future work, we will consider labelling more data and try to design more accurate models to accomplish this task. Besides, because of the sensitivity of sentiment analysis models to the amount of data, transfer learning is also a way to solve this task. Although current transfer learning is still unable to learn from a large number of datasets accurately, we believe this is one of the ways to solve this task in the future.

References

- Bahdanau, D., Cho, K. and Bengio, Y. (2014) *Neural Machine Translation by Jointly Learning to Align and Translate*, arXiv preprint arXiv:1409.0473.
- Blitzer, J., Dredze, M. and Pereira, F. (2007) ‘Biographies, Bollywood, boom-boxes and blenders: domain adaptation for sentiment classification’, *Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics*, pp.440–447.

- Bollegala, D., Weir, D. and Carroll, J. (2012) ‘Cross-domain sentiment classification using a sentiment sensitive thesaurus’, *IEEE Transactions on Knowledge and Data Engineering*, Vol. 25, No. 8, pp.1719–1731.
- Chen, P., Sun, Z., Bing, L. and Yang, W. (2017) ‘Recurrent attention network on memory for aspect sentiment analysis’, *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pp.452–461.
- Conneau, A., Schwenk, H., Barrault, L. and Lecun, Y. (2016) *Very Deep Convolutional Networks for Text Classification*, arXiv preprint arXiv:1606.01781.
- Glorot, X., Bordes, A. and Bengio, Y. (2011) ‘Domain adaptation for large-scale sentiment classification: a deep learning approach’, in Getoor, L. and Scheffer, T. (Eds.) *Proceedings of the 28th International Conference on Machine Learning (ICML-11)*, Bellevue, Washington, USA, 28 June–2 July, pp.513–520.
- He, R., Lee, W.S., Ng, H.T. and Dahlmeier, D. (2018) *Exploiting Document Knowledge for Aspect-Level Sentiment Classification*, arXiv preprint arXiv:1806.04346.
- Hochreiter, S. and Schmidhuber, J. (1997) ‘Long short-term memory’, *Neural Computation*, Vol. 9, No. 8, pp.1735–1780.
- Honnibal, M., Montani, I., van Landeghem, S. and Boyd, A. (2020) ‘spaCy: industrial-strength natural language processing in Python’, *Software Impacts*, Vol. 6, No. 4, p.100263.
- Jiang, L., Yu, M., Zhou, M., Liu, X. and Zhao, T. (2011) ‘Target-dependent Twitter sentiment classification’, *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies – Volume 1*, Association for Computational Linguistics, pp.151–160.
- Kalchbrenner, N., Grefenstette, E. and Blunsom, P. (2014) *A Convolutional Neural Network for Modelling Sentences*, arXiv preprint arXiv:1404.2188.
- Kim, Y. (2014) *Convolutional Neural Networks for Sentence Classification*, arXiv preprint arXiv:1408.5882.
- Kiritchenko, S., Zhu, X., Cherry, C. and Mohammad, S. (2014) ‘NRC-Canada-2014: detecting aspects and sentiment in customer reviews’, *Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014)*, pp.437–442.
- Li, T., Sindhwani, V., Ding, C. and Zhang, Y. (2009) ‘Knowledge transformation for cross-domain sentiment classification’, *Proceedings of the 32nd International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp.716–717.
- Li, Z., Wei, Y., Zhang, Y., Zhang, X. and Li, X. (2019) ‘Exploiting coarse-to-fine task transfer for aspect-level sentiment classification’, *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 33, pp.4253–4260.
- Liu, B. (2012) ‘Sentiment analysis and opinion mining’, *Synthesis Lectures on Human Language Technologies*, Vol. 5, No. 1, pp.1–167.
- Ma, D., Li, S., Zhang, X. and Wang, H. (2017) *Interactive Attention Networks for Aspect-Level Sentiment Classification*, arXiv preprint arXiv:1709.00893.
- Mikolov, T., Karafiát, M., Burget, L., Černocký, J. and Khudanpur, S. (2010) ‘Recurrent neural network based language model’, *Eleventh Annual Conference of the International Speech Communication Association*.
- Mohammad, S.M., Kiritchenko, S. and Zhu, X. (2013) *NRC-Canada: Building the State-of-the-Art in Sentiment Analysis of Tweets*, arXiv preprint arXiv:1308.6242.
- Pan, S.J., Ni, X., Sun, J-T., Yang, Q. and Chen, Z. (2010) ‘Cross-domain sentiment classification via spectral feature alignment’, *Proceedings of the 19th International Conference on World Wide Web*, pp.751–760.
- Pang, B., Lee, L. et al. (2008) ‘Opinion mining and sentiment analysis’, *Foundations and Trends® in Information Retrieval*, Vol. 2, Nos. 1–2, pp.1–135.

- Pennington, J., Socher, R. and Manning, C.D. (2014) ‘Glove: global vectors for word representation’, *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp.1532–1543.
- Socher, R., Perelygin, A., Wu, J., Chuang, J., Manning, C.D., Ng, A.Y. and Potts, C. (2013) ‘Recursive deep models for semantic compositionality over a sentiment treebank’, *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*, pp.1631–1642.
- Tai, K.S., Socher, R. and Manning, C.D. (2015) *Improved Semantic Representations from Tree-Structured Long Short-Term Memory Networks*, arXiv preprint arXiv:1503.00075.
- Tang, D., Qin, B. and Liu, T. (2016) *Aspect Level Sentiment Classification with Deep Memory Network*, arXiv preprint arXiv:1605.08900.
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł. and Polosukhin, I. (2017) ‘Attention is all you need’, *Advances in Neural Information Processing Systems*, Vol. 30, pp.5998–6008.
- Vo, D-T. and Zhang, Y. (2015) ‘Target-dependent Twitter sentiment classification with rich automatic features’, *Twenty-Fourth International Joint Conference on Artificial Intelligence*.
- Wang, W., Pan, S.J., Dahlmeier, D. and Xiao, X. (2016a) *Recursive Neural Conditional Random Fields for Aspect-Based Sentiment Analysis*, arXiv preprint arXiv:1603.06679.
- Wang, Y., Huang, M., Zhu, X. and Zhao, L. (2016b) ‘Attention-based LSTM for aspect-level sentiment classification’, *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pp.606–615.
- Xue, W. and Li, T. (2018) *Aspect Based Sentiment Analysis with Gated Convolutional Networks*, arXiv preprint arXiv:1805.07043.
- Zhang, Y., Barzilay, R. and Jaakkola, T. (2017) ‘Aspect-augmented adversarial networks for domain adaptation’, *Transactions of the Association for Computational Linguistics*, Vol. 5, pp.515–528.