

Publication Date: 10 August 2024

Archs Sci. (2024) Volume 74, Issue S1 Pages 8-13, Paper ID 2024s12.
<https://doi.org/10.62227/as/74s12>

Natural Language Emotion Analysis Model Based on dual attention Convolutional Neural Networks

Hailing Wang^{1,*} and Jian Wang²

¹School of Computer and Information Engineering, Heilongjiang University of Science and Technology, Harbin 150022, China.

²Informationization office, Harbin Engineering University, Harbin 150001, China.

Corresponding authors: Hailing Wang(e-mail:wanghailing@usth.edu.cn).

Abstract Convolutional Neural Networks (CNN) cannot distinguish the correlation between feature words and emotion in input text. Therefore, a dual attention CNN (DACNN) model is proposed that includes two channels: word feature channel and part of speech feature channel. After the feature representation of each channel is obtained, the two features are fused at the global attention level. Finally, the generated text feature representation is used as the input of the full connection layer to obtain the final emotional orientation. The model is verified on MR and SST-1 data sets, and the accuracy is improved by 0.7% and 1% respectively compared with conventional CNN and traditional machine learning methods.

Index Terms CNN, attention mechanism, emotion classification, natural language

I. Introduction

With the rise of social networks such as Facebook and Twitter, individuals are increasingly engaging in commenting on hot events as well as movie reviews described in blogs and on Twitter, among others. Sentiment analysis of these comments provides a better understanding of user behavior and reveals users' tendencies towards certain types of products and attention to current events [1]. With the rapid increase in the amount of information, manual processing is insufficient to accomplish this task [2], [3]. Neural network models in artificial intelligence introduce artificially constructed features such as manually annotated sentiment dictionaries, syntax, and grammar analysis. Although these methods can effectively improve the accuracy of text sentiment analysis, they are time-consuming and laborious because they require a large amount of manually annotated data [4].

With the development of deep learning techniques, researchers have utilized related techniques based on deep neural networks to analyze emotions in texts. For example, recurrent neural networks built on syntactic analysis can discriminate the emotions of movie reviews at different levels [5]. Some researchers believe that deep learning methods are superior at the feature level. Recently, graph convolutional networks (GCNs) have achieved good results in various tasks such as neural machine translation, text classification, semantic role labeling, and natural language inference. For example, [6] introduced GCN into the ABSA task to solve the long-range dependency problem of words in sentences; [6] applied graph attention networks (GAT) to the ABSA task to capture the contribution of contextual nodes to aspectual nodes. GCN

is used to capture the syntactic features of sentences, and two methods are proposed to fuse sequence information and syntactic features to obtain aspect-specific feature representations. In addition, in each dataset, a particular aspect may contain multiple words, and each word has different importance to the particular aspect. To reduce the influence of modifiers such as prepositions (e.g., "of" in "variety of") on aspect-specific features, and to make full use of aspect-specific keywords to learn contextual representations [7].

Aspect-specific sentiment analysis is a fine-grained sentiment analysis task. To reduce manual operations, [8] proposed a semi-automatic approach to construct an ontology database and fuse semantic information to improve the accuracy of sentiment classification; [9] designed a dependency-based rule, and the dictionary was used as an additional reference for sentiment analysis. Most of the methods mainly extract relevant features describing the context and aspects between them and feed them into a classifier for sentiment polarity classification [10]. However, it can be found that the class performance of these methods depends heavily on the quality of knowledge and features and is time-consuming and labor-intensive.

RNN-based approaches preserve important contextual information by capturing sequential information of the comment data [11]. For example, based on TC-LSTM models and attention mechanisms, different weights are assigned to context words by computing the semantic relevance of specific aspects and context words to enhance the influence of aspects on sentiment classification [12]; by modeling multiple aspects of the same sentence, the noise introduced by attention mecha-

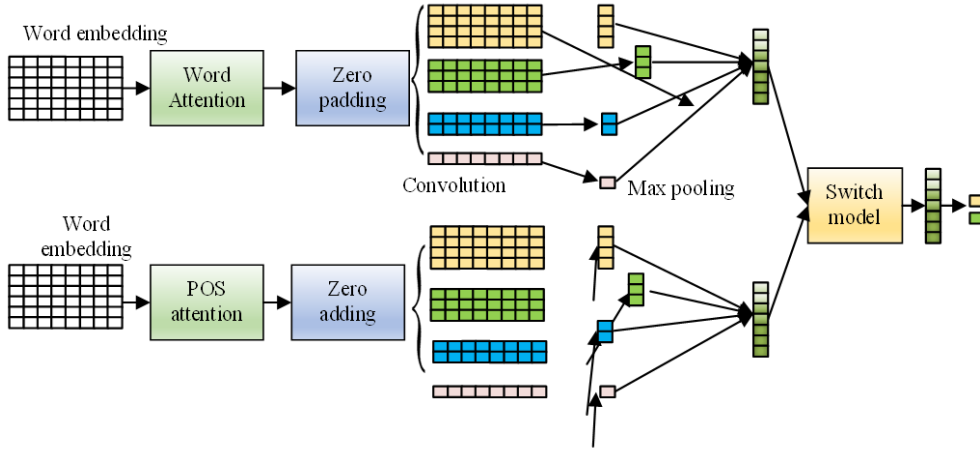


Figure 1: Dual Attention CNN Model

nisms in ABSA tasks can be reduced [13]. Another part of the work involves building CNN-based models. For example, [14] proposed a model based on a gating mechanism and a CNN that uses convolutional kernels to compute multigranularity grammatical features and uses a gating mechanism to update irrelevant features.

In this paper, a structure based on CNN and attention models is adopted to avoid the method of relying on the artificial construction of features.

II. Word Attention CNN

This paper proposes a dual attention CNN model, as shown in Figure 1. The model includes two channels: a word feature channel and a part-of-speech (POS) feature channel. After obtaining the feature representation for each channel, the two features are fused at the global attention level [15]. Finally, the generated text feature representation is used as the input to the fully connected layer to determine the final sentiment polarity.

A. Local Attention CNN

In this paper, text features are extracted from both word and part-of-speech features. The mechanism for the two channels is identical, both adopting a local attention convolutional neural network model. Therefore, this section introduces the model structure primarily from the word perspective [16].

As illustrated in Figure 1, the local attention convolutional neural network model proposed in this paper is mainly divided into five layers: Word Embedding Layer, Word Attention Layer, Padding Layer, Convolution Layer, and Pooling Layer.

Word Embedding Layer: The text is represented by word feature mapping through the model as $S = \{w_1, w_2, \dots, w_{n-1}, w_n\}$, where w_i represents the i -th word in the sequence. A one-hot model is used for text feature representation [17]. The word or phrase is mapped into a d -dimensional vector through the word embedding layer.

$$x_i = w_i e, \quad x_i \in R^d \quad (1)$$

In Eq. (1), e represents the word vector matrix, and the matrix after word embedding is $R^{d \times |v|}$, where $|v|$ represents the size of the entire vocabulary of the dataset.

Word Attention Layer: After the word embedding layer, the word attention layer is added, whose primary function is to distinguish the information represented by the text features, highlight key word features, and allow the model to selectively focus on important information. In this paper, the n -gram language model is used to express the words in the context of the text as features of the word. Here, D is used to represent the context range of a central word, and the size of the word window is $L = [p_i - D, p_i + D = 1.2D]$. This serves as a new expression for the word. $W_{att} \in R^{L \times d}$ represents the sliding window matrix parameter. Each word's eigenvalue weight α_i is calculated in this step to indicate the importance of words in the text, as shown in Figure 2.

$$\begin{cases} \mathbf{X}_i^{att} = \{\mathbf{x}_{i-D}, \dots, \mathbf{x}_i, \dots, \mathbf{x}_{i+D}\} \\ \mathbf{X}^{att} = \{\mathbf{X}_1^{att}, \dots, \mathbf{X}_n^{att}\} \end{cases} \quad (2)$$

$$\alpha_i = h(\mathbf{X}_i^{att} \otimes \mathbf{W}_{att} + b_{att}), \quad (3)$$

where n is the length of the sentence in the text, α represents the importance value of a word or phrase in a sentence. Each word feature representation x_i is multiplied by α to output a new feature representation X_{att} .

$$\begin{aligned} \mathbf{X}_i^{att} &= \alpha_i \mathbf{x}_i, \quad \mathbf{X}_i^{att} \in R^d \\ \mathbf{X}^{att} &= \{\mathbf{X}_1^{att}, \dots, \mathbf{X}_i^{att}, \dots, \mathbf{X}_n^{att}\}. \end{aligned} \quad (4)$$

Padding Layer: Traditional sentiment analysis methods for natural language processing typically use unigram, bigram, and trigram models to extract feature representations for sentiment analysis [17]. These methods have been used in sentiment analysis and have achieved good experimental results [18]. In this paper, when the convolution kernel size is 3, the context range for each headword in the text is 1, following the Markov principle. When the convolution kernel size is 5, the

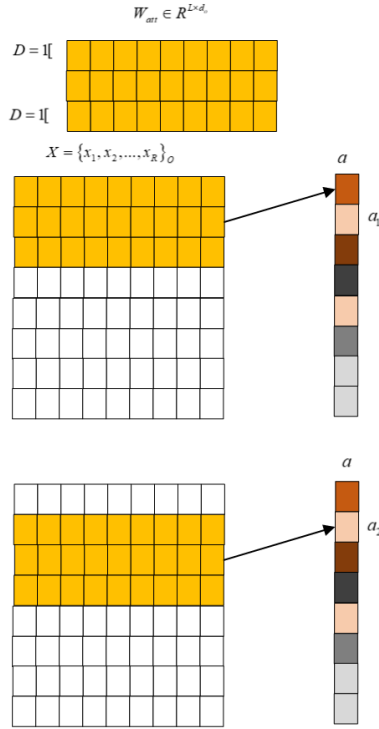


Figure 2: Word Attention Layer

context range expands to 2. However, this leads to insufficient text information extraction during model training. During the convolution operation, the first word cannot be retrieved from above, and the last word cannot be retrieved from below [19]. To address this, two padding methods are proposed:

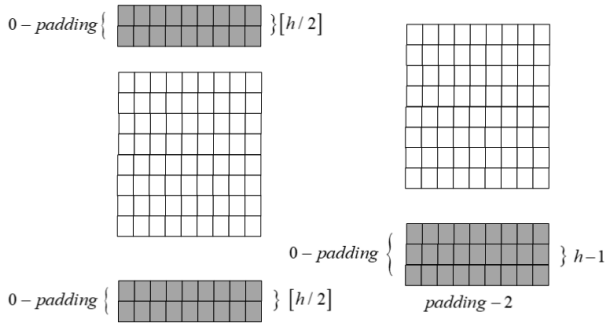


Figure 3: Two Padding Methods

- (1) **Padding-1:** Fill the beginning and end of the input to the convolution layer with a zero vector. Since the information of the first and last parts is missing in half of the window, the size of the padding is chosen to ensure that the first word exists above and the last word exists below.

$$X = \underbrace{X_0 \oplus X_0 \oplus \dots \oplus X_{att}}_{[h/2]} \oplus \dots \oplus \underbrace{X_0 \oplus X_0}_{[h/2]}. \quad (5)$$

- (2) **Padding-2:** Complete the end of the convolution layer input with a zero vector of size $h - 1$.

$$X = X_{att} \oplus \dots \oplus \underbrace{X_0 \oplus X_0}_{h-1}. \quad (6)$$

Here, $\oplus$ represents the concatenation operation.

B. Global Attention Fusion Layer

Suppose the word feature representation V_W and the part-of-speech feature representation V_{POS} of the text are obtained in the dual channel. On the one hand, the contribution of word feature representation and part-of-speech feature representation to emotional feature representation is different, and the weight given in feature fusion is also different [20]. On the other hand, the traditional concatenation method is used for feature fusion, which doubles the dimension of the text feature representation after fusion. Therefore, a global attention mechanism is proposed to fuse different features and assign higher weights to important features.

$$u = \tanh(\mathbf{W}_{sm}[v_w, v_{pos}] + \mathbf{b}_{sm}), \quad u \in [u_w, u_{pos}]. \quad (7)$$

$$a_t = \frac{\exp(u_t)}{\sum_t \exp(u_t)}, \quad a_t \in [a_w, a_{pos}]. \quad (8)$$

$$s = a_w v_w + a_{pos} v_{pos}. \quad (9)$$

The feature representation $[v_w, v_{pos}]$ obtained from local attention is used as the input of the Multi-Layer Perceptron (MLP) to get the input u , where $\mathbf{W}_{sm}$ is the weight of the MLP. The input u is normalized by the softmax function to obtain the weight of each feature. Finally, the feature representation $[v_w, v_{pos}]$ is multiplied by the corresponding weight and summed to get the feature representation s of the text.

III. Experimental Results

Two sets of experiments are designed to verify the effectiveness of the proposed method. Experiment 1 compares the performance of the WFCNN model proposed in this paper with the CNN model based on word2vec training (labeled W2VCNN) proposed by Kim and the NBSVM model proposed by Sida Wang. In Experiment 2, the emotion sequence features extracted by this method are added to the W2VCNN model and the NBSVM model to verify the effectiveness of this method in extracting text emotion features.

A. Experimental Data Preprocessing and Model Parameter Setting

The dictionary resources used in this experiment are composed of MR and SST-1 datasets. For data preprocessing, the ICTCLAS word segmentation tool is used to segment and label the experimental dataset. The word vectors adopt the Skip-gram model of Google's open-source word2vec, and are

Adjustable parameter	value
Convolution kernel function	Corrected linear function
Filter sliding window size h	3,4,5,6
Number of filters m	150
Update parameter scale randomly	1.0
Number of training iterations	100

Table 1: Parameter setting

Model	Precision+, Precision−	Recall+, Recall−	F−score+, F−score−
WFCNN	0.7805, 0.7502	0.7901, 0.7382	0.7855, 0.7454
W2VCNN	0.7872, 0.8457	0.8865, 0.7194	0.8332, 0.7759
NBSVM	0.7172, 0.7553	0.8281, 0.6194	0.7688, 0.6792

Table 2: Comparison of emotion classification performance of three models (COAE2014 data set)

generated by training with a 20-million microblog corpus. The word vector dimension is 50, including 330,000 words, with a vocabulary coverage rate of 90.08% on the experimental dataset. The parameter settings of our model are shown in Table 1.

In the NBSVM model, this paper uses Unigram and Bigram language models to construct text feature vectors.

B. Model comparison experiment

The first group of experiments compares the performance of the WFCNN model proposed in this paper with other existing models. Table 2 lists the comparison results of the experiment on the COAE2014 data set.

From the experimental results in Table 2, we can see that the WF-CNN model proposed in this paper has a certain gap in performance with the W2VCNN based on word2vec's word vector. WFCNN model is based on dictionary resources to extract the general rule sequence that expresses the emotional orientation of text, and has strong model generalization ability. Compared with the traditional NBSVM based on domain knowledge, WFCNN not only has advantages in model training, but also has great advantages in performance. According to the experimental results in Table 2, the WFCNN of positive emotion recognition is 1.69% higher than that of NBSVM, while the WFCNN of negative emotion recognition is 6.51% higher than that of NBSVM.

C. Comparison of feature fusion experiments

In the second experiment, in order to verify the effectiveness of the WFCNN model proposed in this paper in extracting text emotional orientation feature sequence, the text feature vectors generated by the WFCNN model are fused with the feature vectors of the W2VCNN model and the NBSVM model respectively, and the model performance changes before and after the feature fusion are compared. Table 3 lists the comparison of experimental results before and after feature fusion.

By comparing the experiments before and after the model feature fusion, it can be found that the text emotional orientation sequence features generated by WFCNN model can effectively improve the effect of text emotional classification. For the W2VCNN model, after integrating the emotional

tendency sequence features generated by the WFCNN model, the F-score of positive emotion recognition was improved by 0.97%, while the F-score of negative emotion recognition was increased by 1.59%. For the NBSVM model based on domain knowledge, the classification performance has been greatly improved after the introduction of the emotional orientation sequence features generated by the WFCNN model, with 4.97% and 7.52% improvement in the F-score of positive and negative emotion recognition, respectively.

D. Results Analysis

Here, several concrete examples are used to analyze the effectiveness of WFCNN model in introducing text emotion sequence features. Table 4 shows the representative data machine classification results selected from the test data set.

As shown in Table 4 that the classification results of the sample data. For example 1, "Samsung's mobile phone is not used to", and example 2, "Jewelry is a good investment, especially the appreciation space of jadeite is n times that of real estate". Because this kind of example expression is relatively common, there are similar examples in the training set and test set. The three methods have a certain degree of coverage for this kind of example, so they have made correct emotional classification.

For example 3, "adding insurance is the best way to manage risks scientifically and transfer losses." and example 4, "buying insurance means that one trouble will be followed by another trouble." Because the emotional words in this kind of example do not work alone, but express a certain logical relationship through the sequence of words. Such samples are incorrectly classified in the CNN based on word vectors, while the CNN based on sequence features can be processed correctly. After integrating these sequence features, the method based on word vector can correct the previous errors by classification. For example 5, "Does Huawei only know technology?" All methods are not classified correctly. Because of the rhetorical and ironic expressions in this example, the key word "skill" is judged as a positive example whether it is based on word vector or dictionary. Therefore, this sample is not correctly classified in the three methods.

It can be seen that the WFCNN model can accurately identify the emotional tendency of the text by extracting

Model	Precision+, Precision—	Recall+, Recall—	F—score+, F—score—
W2VCNN	0.7875,0.8439	0.8869,0.7193	0.8335,0.7761
W2VCNN+WFCNN	0.8002,0.8537	0.8928,0.7402	0.8452,0.7933
NBSVM	0.7176,0.7538	0.8271,0.6183	0.7685,0.6796
Our model	0.7721,0.8229	0.8721,0.6868	0.8185,0.7547

Table 3: The influence of emotion sequence characteristics on emotion classification performance (COAE2014 data set)

Data sample	WFCNN	W2VCNN	WFCNN+W2VCNN	Manual marking
Star's mobile phone is a little strange!	1	1	1	1
Jewelry is a good investment, especially the appreciation space of jadeite is n times of real estate.	-1	-1	-1	-1
Joining insurance is the best way to scientifically manage risks and transfer losses.	-1	1	1	-1
Buying insurance means that one trouble will be followed by another.	1	-1	1	1
Does Huawei only know technology.	-1	-1	-1	1
Samsung, you have to go against the sky and bend like that. Emma is too arrogant, and the mobile phone is also yoga. overbearing.	1	-1	-1	-1
After I found that I used Samsung mobile phone, my microblog was full of typos! Can't bear the bird! Be careful, be careful!	1	1	-1	-1

Table 4: Example of test data experiment results (COAE2014 data set)

the sequence features containing emotional words, such as "scientific management risk", "the best way", "some troubles", etc. However, W2VCNN model can not extract the sequence features expressing emotional orientation in the text because it only uses the word vector of the grammatical and semantic information between words, so it classifies the text incorrectly. After adding the emotional sequence feature of WFCNN model to W2VCNN model, the model can accurately identify the emotional orientation of text. In addition, it can be found from the test samples that neither WFCNN nor W2VCNN model can recognize the emotional tendency of irony sentence patterns.

4.5 Weight change of attention fusion layer
It can be seen from Figure 4 that the performance of DACNN-N model on multiple data sets is slightly lower than that of DACNN1 and DACNN2 According to Figure 4, the performance of DACNN-N is improved compared with the original CNN, but compared with DACNN1 and DACNN2, the performance is poor. During the training process, the model further extracts the features of the text in the convolution layer, adds a filling layer to prevent the problem of information loss, and improves the effect of the experiment. Figure 6 shows that the advantages and disadvantages of different filling methods are uncertain, and their effects are different in different target data Which is better depends on the specific goal.

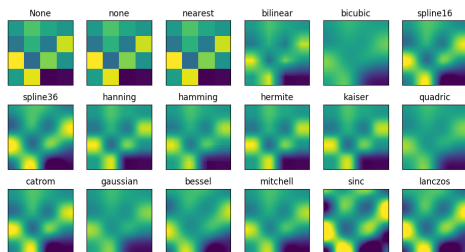


Figure 4: Weight change of attention fusion layer

E. Word attention visualization

In order to help better understand the impact of word attention on the model, this part visualizes word attention for feature

extraction As shown in Figure 6 and 6 The visualization experimental data are two texts randomly selected from the two experimental data sets in this paper Figure 5 shows the text data of positive emotions and Figure 6 takes the data text of negative emotions The weight value of the word feature representation is obtained at the word attention level The text is labeled with different color depths, and the weight of color depth in the experiment is relatively large A light color indicates a small weight In order to facilitate visual viewing, other features with lower weight words are not marked As shown in Figure 5 is the randomly selected data text in SST-1, we can see that word attention really has the function of focusing on keywords.

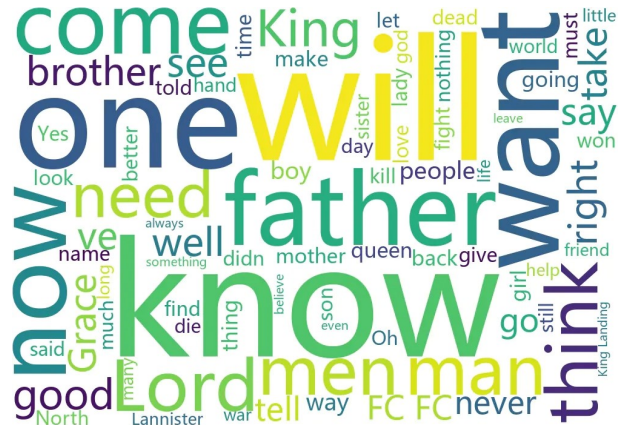


Figure 5: Shows text data with positive emotions

IV. Conclusion

Through the research of emotion analysis, this paper proposes a convolutional neural network model based on double attention, and the experiment has achieved good results First, the model extracts features from word and part of speech channels Secondly, the method of local attention mechanism is used to extract feature words in the training process of the model, which improves the ability of word feature selection of the model Finally, the visual analysis of the attention mechanism layer fully proves that the word attention mechanism added to the model in this paper is helpful to the emotional analysis

