Research on Text Multi-Feature Fusion Algorithm Based on AM-CNN

WANG Wensong a and SUN Xiange b

College of Electronic & Information, Yangtze University, Jingzhou Hubei 434023, China

a201971324@yangtezu.edu.cn; b100854@yangtezu.edu.cn

Abstract: How to know the changes of students' emotions in real time in online education has always been one of the important issues concerned by education departments and teachers. Making use of the special functions of global, part of speech and positional attention mechanisms in text processing, a deep learning network model based on the combination of multiple attention mechanisms and convolution neural networks is proposed. Firstly, the blending characteristics between the types of attention mechanism and CNN are analyzed by using the standard ChnSentiCorp_htl_all data set to clarify the effectiveness of the combination of the three attention mechanisms and CNN. Then the model is applied to the analysis of the evaluation text of the course "big data Technology principles and applications" on the MOOC of Chinese universities, and it is verified that the evaluation indexes of this model are better than the existing conventional models.

1 Introduction

With the development of Internet technology, online education has become an important way for students to enrich themselves in their spare time. The change of students' emotions in the learning process is the focus of researchers, and the commonly used method is to use the textual information of students' evaluation of courses to analyze. At present, the methods for text analysis mainly include support vector machines [1], naive Bayes [2], Kmeans clustering [3] and neural networks [4]. In the traditional method, the data form has high dimensional sparseness, the representation ability is not strong, and the features affect each other, so it is easy to cause the model's convergence time to be too long and the text analysis [5] to be less effective.

With the development of deep learning and computer hardware conditions, more and more neural networks are used to deal with various natural language problems. For example, Kim [6] et al. introduced Convolutional Neural Network (CNN) to text classification problems in natural language for the first time in 2014, and performed convolution and pooling operations on sentence-encoded vectors. Because CNN considers It is a local problem, and the effect is not very obvious compared to the traditional RNN network that solves natural language problems; Lai [7] et al. proposed in 2015 to use CNN and RNN to combine to build a recurrent convolutional neural network (Recurrent Convolutional Neural Networks, RCNN), Send the features processed by RNN to CNN to achieve text classification. During the implementation process, it was found that the text classification results can be obtained directly during RNN processing; Conneau [8] et al. proposed in 2016 to use a deeper CNN to solve the text classification problem, Found the network level that makes the text classification model the best.

In recent years, more and more scholars have begun to use the attention mechanism to deal with
problems. For example, Vaswani [9] et al. applied the attention mechanism to natural language problems in 2017. Compared with the traditional neural network model, the attention mechanism can be parallelized. The research in this paper confirms that attention is paid to machine translation tasks. The efficiency of the mechanism is significantly improved compared to traditional methods; Letarte [10] and others used the self-attention mechanism to solve the problem of text classification in 2018, which is significantly improved compared to the RNN network.

Some scholars have proposed whether the combination of the two can be used to achieve better results. This idea was first applied to the field of image processing to achieve relatively ideal results. In accordance with the solution of the attention mechanism in the field of image processing, try to use it in natural language processing problems. Therefore, in 2016, Yin W [11] et al. used this type of model for the first time to process text information, adding an attention mechanism to the convolutional layer, the pooling layer, and the two at the same time to pass the relationship between words and sentences. The attention mechanism performs weighting processing, which has achieved a relatively obvious improvement; Wang L [12] and others used the combination of multi-head attention mechanism and CNN to solve the text classification problem in 2018, and used specific relationship extraction to perform weight processing on word vectors. Because part of speech and positional relations are not considered, the error rate is increased when dealing with implicit relations, resulting in a slight improvement in the original results of the model.

At present, the combination of attention mechanism and network model has become a major trend. Using attention mechanism can extract more useful features, and then can represent information well. This paper attempts to solve the problem of sentiment analysis of text by combining various Attention mechanisms (AM) and CNN to build an AM-CNN model for the text information of course evaluation. In the process of data set preprocessing, the attention mechanism is used to process the global information, part of speech information and location information in the text information into text vector processing, and further feature extraction and fusion are completed through the CNN model, and finally the emotional category of the text is obtained. Compared with traditional models, it has better results.

2 AM-CNN Algorithm Model

2.1 Analysis and Selection of Attention Mechanism

The attention mechanism model in deep learning is essentially similar to human selective attention. The core principle is to find out the information to be paid attention to from a lot of information. In previous studies, the use of self-attention mechanism to process word vectors can obtain the relationship of words with respect to a certain sequence; in Transformer, researchers use self-attention to process word vectors and divide Q, K, and V during processing. The multi-head attention mechanism is obtained, where Q, K, and V represent the initialization vectors of “query”, “key” and “value” respectively, and then the corresponding weight word vector is obtained. For a sentence, the overall analysis is important, but some partial analysis can also be a good supplement. The self-attention mechanism only considers the influence of other words in the text sequence on the analyzed words when processing the text. In linguistics, the different positions of the words in a sentence may express different meanings, so the position of the words is more important for the understanding of the sentence; at the same time, in the process of analyzing the sentence, if you can know the emotional trend of certain keywords, then This sentence has important implications. According to the characteristics of the text language and previous studies, this article attempts to analyze the text using global attention, position attention and part-of-speech attention mechanisms.

In the part-of-speech attention model, the text information is processed using the emotional dictionary in SnowNLP [13] to obtain the score of the emotional word, and then the similarity between the word vector and the emotional word is used to obtain the weight matrix of the word vector, and finally the original the word vector matrix convolution operation obtains the part-of-speech vector matrix.

In the positional attention model, the positional encoding method of the Transformer model is used
for processing. Assuming an input sequence, the position of a word in the sequence is \( t \), \( \mathbf{p}_t \in \mathbb{R}^d \) represents the vector at position \( t \), and \( d \) is the dimension of the word. \( f: \mathbb{N} \rightarrow \mathbb{R}^d \) is a function to generate a position vector \( \mathbf{p}_t \), defined as:

\[
\mathbf{p}_t^i = f(t)^i = \begin{cases} 
\sin(w_k \cdot t), & \text{if } i = 2k \\
\cos(w_k \cdot t), & \text{if } i = 2k + 1
\end{cases}
\]

among them: \( w_k = \frac{1}{10000^{2k/d}} \)

Suppose an input sequence \( w = (w_0, ..., w_{L-1}) \) of length \( L \), the dimension of the word is 4, and the word vector corresponding to the word \( w_t \) is \( \mathbf{v}_w \), \( t \in [0, L-1] \). Then the position vector corresponding to \( w_t \) can be calculated according to formula (1):

\[
\mathbf{p}_t = \left[ \sin(w_0 \cdot t), \cos(w_0 \cdot t), \sin(w_1 \cdot t), \cos(w_1 \cdot t) \right] \\
= \left[ \sin\left(\frac{t}{10000^0}\right), \cos\left(\frac{t}{10000^0}\right), \sin\left(\frac{t}{10000^2}\right), \cos\left(\frac{t}{10000^2}\right) \right] \\
= \left[ \sin(t), \cos(t), \sin\left(\frac{t}{100}\right), \cos\left(\frac{t}{100}\right) \right]
\]

The word \( w_t \) finally means: \( w_t = \mathbf{v}_w + \mathbf{p}_t \).

In the global attention model, if you want to get the form of the word vector in the sentence after attention transformation, you need to initialize three parameter matrices \( W^Q, W^K, W^V \), and divide the word vector matrix with the three parameters. The matrix is convolved to obtain the values of \( Q, K, \) and \( V \), and then the global vector matrix transformed by the attention mechanism can be obtained by using formula (2).

\[
\text{Attention}(Q, K, V) = \text{Softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V
\]

Among them, \( Q, K, \) and \( V \) represent "query", "key" and "value" respectively; \( d_k \) is the scaling factor, and the value is equal to the dimension of \( K \). As a larger value of \( d_k \) makes the fraction value extremely small, the gradient disappears, in order to compensate this influence takes the form of \( \frac{1}{\sqrt{d_k}} \).

2.2 Model Building

The AM-CNN model uses three types of attention mechanisms to generate the corresponding training set of the initial text vector, and uses the word vector to transform the text vector to process the part-of-speech attention mechanism. The method used is to use SnowNLP to analyze each text in the text. Calculate the scores of each word to form a part-of-speech weight matrix, and convolve with the text vector to obtain the part-of-speech vector matrix \( x_{\text{Set1}} \): in the same way, use the position coding information to perform position weight processing on each word to obtain the corresponding position weight vector, and convolve with the text vector to obtain the position vector matrix \( x_{\text{Set2}} \); Use the self-attention mechanism to obtain the weight information of the impact of each word on the sentence to obtain the global weight vector, and convolve with the text vector to obtain the global vector matrix \( x_{\text{Set3}} \): The three pre-processed vector matrices are sent to the convolutional neural network respectively, and the vector matrices under different attention mechanisms are extracted through the convolution layer, and then the data matrix is reduced by the pooling layer; adding a merge layer integrate the data after three kinds of vector matrix pooling with different angles; send the integrated data to the fully connected layer and the output layer for processing, and finally get the classification result. The flowchart is shown in Figure 1.

2.3 Basic Judgment Basis for Model Training

The test set data is in the two categories under the model, TP and TN respectively represent the number of samples predicted to be 0 and 1 when the real sample is 0, and FP and FN are respectively the real sample of 1 In this case, the number of samples predicted to be 0 and 1. The conventional discrimination basis of the two-class model is the accuracy rate, precision rate, recall rate and F1 measurement value.
3. Experimental Processing and Analysis

The process of this experiment was carried out under the Windows operating system, using the TensorFlow2.0 framework to build the network model, where the computer memory size is 8G, the python version uses Python 3.7, and the integrated development environment is in Pycharm2019.1 on.

![AM-CNN model frame diagram](image)

**Figure 1.** AM-CNN model frame diagram

3.1 Analysis and Comparison of Multiple Attention Mechanisms

In order to verify the influence of different attention mechanisms on the final results, the fusion analysis of attention mechanisms was carried out, and the ChnSentiCorp_htl_all standard data set was processed by combining different attention combinations and CNN, and the analysis results in Table 1 were obtained.

| Model Name | Accuracy | Recall | F1   |
|------------|----------|--------|------|
| CNN+ Global attention | 0.8315 | 0.8267 | 0.8291 |
| CNN+ Speech attention | 0.8345 | 0.8403 | 0.8374 |
| CNN+ Position attention | 0.8315 | 0.8456 | 0.8357 |
| CNN+ Global & Position | 0.8415 | 0.8512 | 0.8463 |
| CNN+ Global & Speech | 0.8526 | 0.8621 | 0.8573 |
| CNN+ Position & Speech | 0.8505 | 0.8556 | 0.8530 |
| AM-CNN | 0.8912 | 0.8953 | 0.8932 |

It can be seen from Table 1 that under the influence of a single attention model, it is the part-of-speech attention mechanism that has the greatest impact on the final sentiment classification results, which shows that in the short text data, the part-of-speech is compared with the position information and the global information weight. Bigger. Under the influence of the two attention models, the most important influence on the final sentiment analysis results is the combination of part of speech & global attention mechanism, which shows that on the basis of part of speech, global information weights can extract features in more detail. But compared to AM-CNN, there is no obvious effect of three combinations of single and two, which shows that the feature extraction of text from three different angles can better express text information.
3.2 Analysis of Course Evaluation Information

3.2.1 Data Preprocessing. The original data used in the experimental test in this paper is to use the Scrapy framework to crawl the evaluation information text and the scoring of the text from the "Principles and Application of Big Data Technology" course on the MOOC website of Chinese University. Label processing is performed according to the scoring standard of the evaluation text. The sentiment of the evaluation text greater than Samsung is positive, the sentiment of the evaluation text equal to Samsung is neutral and the sentiment of the evaluation text less than Samsung is negative. The corresponding numbers are expressed as: 0 for negative, 1 for Neutral and 2 are positive. The original data text needs to undergo preprocessing operations. The steps are: use jieba for word segmentation of the sentence text; apply stop word processing to the segmentation result; use Word2Vec for word vector conversion for the data, and use the text data of the file itself for dictionary training. After preprocessing, each piece of information text is saved in the form of a two-tuple, and its format is: <text data matrix, emotional label>. Finally, there are 10,000 data sets, and the ratio of corresponding labels in the data set is 0:1:2=10%:75%:15%.

3.2.2 Evaluation and Analysis of AM-CNN Model in Course Evaluation

| mood    | Accuracy | Recall | F1   |
|---------|----------|--------|------|
| negative| 0.88     | 0.87   | 0.88 |
| neutral | 0.90     | 0.90   | 0.90 |
| positive| 0.85     | 0.82   | 0.83 |
| Accuracy|          |        | 0.87 |
| Macro average | 0.87 | 0.86 | 0.87 |

Table 2. The Index of AM-CNN Model in Test set

![Figure 2. Confusion matrix of test set](image)

The text of the course evaluation text data set is vectorized, and the obtained data is divided into a test set and a training set, which are divided into equal proportions according to the proportion of labels. Send the training set to the built AM-CNN model for training until the model converges. The convergence model is verified through the test data set, and the results of the relevant measurement indicators are shown in Table 2 below. The confusion matrix for the prediction result of the course evaluation text information is shown in Figure 2.

It can be seen from the confusion matrix of the model that the neutral text recognition rate for the evaluation information is relatively high. This is because the model has more neutral text data during the training process and can extract relevant features well, while the other two There are fewer kinds of data, and the recognition rate is relatively low.

3.3 Sentiment Analysis of Text Sample Set of Customized Course Evaluation

In order to verify the emotion classification model under the custom curriculum evaluation text sample set, we choose different models to analyze the effect of am-cnn model under the custom data set. After training and testing the data set, the result data in Table 3 is obtained.
Table 3. Indicators in the test set under different models

| Model Name | Accuracy | Recall | F1    |
|------------|----------|--------|-------|
| SVM        | 0.8125   | 0.8119 | 0.8121|
| CNN        | 0.8243   | 0.8227 | 0.8235|
| Text CNN   | 0.8559   | 0.8523 | 0.8541|
| Self-Att   | 0.8739   | 0.8668 | 0.8703|
| TextGCN    | 0.8795   | 0.8526 | 0.8658|
| XLNet      | 0.8515   | 0.8912 | 0.8709|
| ALBERT     | 0.8416   | 0.8625 | 0.8519|
| AM-CNN     | 0.8914   | 0.8798 | 0.8856|

From the model test results in Table 3, it can be seen that the AM-CNN algorithm model is better than other models in the process of processing course evaluation text data. We know that ALBERT is the latest improved BERT model proposed in 2020. A good improvement has been made in the preprocessing process, but the effect of the short text classification in this article is not very obvious. Compared with other models, it is found that there is no effect of TextCNN proposed in 2014 and TextGCN proposed in 2019. Obviously, it may be due to the limited amount of information in short texts. Excessive model results lead to poor model over-fitting noise prediction. As mentioned in the review, AM-CNN is more effective in processing course evaluation text data.

In order to verify whether the AM-CNN model can be used in production, we use the AM-CNN model to predict the label of the text through the custom text information. The results are consistent with human ideas. The text information is: 1) This course is too long-winded; 2) The teacher said a lot of useless things; 3) It’s okay, acceptable; 4) Lessons learnt a lot; 5) The trainer's thoughts are very clear, and I can understand it quickly.

The above five pieces of text information are predicted by the model, and the results are shown in Table 4.

| Text       | (Probability) | (Probability) | (Probability) | Corresponding number label | Corresponding to Chinese label |
|------------|---------------|---------------|---------------|----------------------------|-------------------------------|
| Text 1     | 0.99719113    | 0.00280396    | 0.00000492    | 0                          | negative                     |
| Text 2     | 0.8948034     | 0.00823929    | 0.09695735    | 0                          | negative                     |
| Text 3     | 0.00000004    | 0.99999999    | 0.00000007    | 1                          | neutral                      |
| Text 4     | 0.00000047    | 0.99999975    | 0.00000205    | 1                          | neutral                      |
| Text 5     | 0.00001114    | 0.00012788    | 0.999861      | 2                          | positive                     |

4. Summary
This article is to conduct research on the evaluation text of online courses, and proposes a multi-feature fusion text analysis model AM-CNN, which uses multiple attention mechanisms to preprocess the original data and then perform feature extraction and fusion through CNN. Make the analysis result more perfect. This study also verified the practicability of the model on the public Chinese ChnSentiCorp_hlt_all data set and English SemEval data set, and compared it with traditional methods. Through experimental verification, it is obtained that the attention mechanism can improve the problem of CNN's insufficient processing of global information. Both Chinese and English data collection have achieved good results, and it provides an effective way to process text information.

During the research process, it was found that the effect of Chinese processing was significantly lower than the English data set. The next step will focus on how to optimize the model to improve the effect on the Chinese data set and how to deal with data sets with uneven samples.

References
[1] Peng Danlei, Gu Lize, Sun Bin. Research on Sentiment Analysis of Product Reviews Based on Two Models of SVM and LSTM[J]. Software, 2019,40(01):41-45.
[2] Wang Binjing. Research on Sentiment Analysis of Weibo Text Based on Naive Bayes Classification Algorithm[J]. China New Telecommunications, 2019, 21(08):114-115.

[3] Liu Jianghua. A Text Retrieval Method and Validity Verification Based on Kmeans Clustering Algorithm and LDA Topic Model[J]. Information Science, 2017, 35(02):16-21+26.

[4] Li Weijiang, Yi Jing. Weibo Text Sentiment Classification Based on Extended Feature Matrix and Two-Layer Convolutional Neural Network[J]. Computer Applications and Software, 2019, 36(12):150-155.

[5] Zhang Weiwei, Hu Yaqi, Zhai Guangyu, et al. Academic Abstract Clustering Method Based on LDA Model and Doc2vec[J]. Computer Engineering and Applications, 2020, 56(06):180-185.

[6] Kim Y. Convolutional neural networks for sentence classification[C] //Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP-P02014), Doha, Qatar: ACL Press, 2014:1746-1751.

[7] Lai S, Xu L, Liu K, et al. Recurrent convolutional neural networks for text classification[J]. The Association for the Advance of Artificial Intelligence, AAAI, 2015, 33:2267-2273.

[8] Conneau A, Schwenk H, Barrault, et al. Very deep convolutional networks for text classification[C] // Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics, EACL, 2017:1107-1116.

[9] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need[C]// Proceedings of 2017 Advances in neural information processing systems. Long Beach, California, USA: Curran Associates, Inc. 2017.

[10] Letarte G, Paradis F, Giguere P, et al. Importance of self-attention for sentiment analysis[C]//Proceedings of the 2018 EMNLP Workshop BlackboxNLP: Analyzing and Interpreting Neural Networks for NLP. Brussels, Belgium: [s.n.], 2018.

[11] Yin W, Schutze H, Xiang B, et al. ABCNN: attention-based convolutional neural network for modeling sentence pairs[J]. Transactions of the Association for Computational Linguistics, 2016, 4(11):259-272.

[12] Wang Linlin, Cao Zhu, de Melo G, et al. Relation classification via multi-level attention cnns[C]. Proc of the 54th Annual Meeting of the Association for Computational Linguistics. Association for Computational Linguistics. Stroudsburg, PA: ACL, 2016:1298-1307.

[13] Caixia Chen, Jue Chen, Chun Shi. Research on Credit Evaluation Model of Online Store Based on SnowNLP[C]// Proceedings of the Third International Conference on Energy and Environmental Research Progress (ICAEER 2018).