ISSN: 2472-3703 DOI: 10.6911/WSRJ.202211_8(11).0005

Sentiment Analysis on Online Learning Comments via CNN-BiLSTM with Attention

Shuqin Huang[1, a], Xiaoyan Shi[1, b], Sulin Wei[1, c], Yong Xu[1, d] and Hengna Wang[1, e]

[1]School of Management Science and Engineering, Anhui University of Finance &Economics, Bengbu, China

[a]393054429@qq.com, [b] xiaxiaoxi2009@163.com, [c]wei_sulin@163.com, [d]uxyong@sina.com, [e]928290228@qq.com

Abstract

Sentiment analysis is an important part of natural language processing. Students' learning status can be evaluated by analyzing their sentiment in online course comments. The CNN-BiLSTM with attention model was established to evaluate the students' emotional state. This model can avoid the problem that convolutional neural networks ignore the contextual meaning of word, traditional recurrent neural network is easy to cause gradient disappearance and LSTMs can't prioritize individual sections of the sentence that are relevant. The experiments on real online course comments showed that our proposed model outperforms three baselines in accuracy.

Keywords

CNN; BiLSTM; Attention mechanism; Text sentiment analysis; Online course comments.

1. INTRODUCTION

With the development of online teaching and MOOCs, many "Internet + education" learning platforms have emerged at home and abroad, along with the emergence of more student learning data. However, traditional learning analysis techniques usually focus on numerical analysis, such as explicit behavioral data, the duration of learners' browsing courseware and the times of submitting/not submitting homework in online learning system, while less consideration is given to implicit behavioral information such as learners' attitude and state. Emotion is an important factor affecting the learning effect of students. Mining the students' sentiment contained in the learning data can directly master the learning state of students. Positive emotion can motivate students to learn, while negative emotion can frustrate students' interest in learning. How students' emotions change will directly affect the quality of a student's learning. If students' negative emotions in learning are found in time and targeted solutions are taken, students will gradually turn to a positive state, which is conducive to improving the learning and teaching effect.

2. RELATED WORKS

With the rise of educational data mining, scholars begin to study the emotion in educational learning behavior. As far as the data object of sentiment analysis is concerned, it mainly includes sentiment classification, recognition and analysis research based on text, image, audio and video. Text is the most typical and explicit instance source in unstructured data. And it is also one of the most studied objects in sentiment analysis. For learning experience texts, Li[1] proposed a learner's emotion analysis model combining emotion dictionary and machine

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202211_8(11).0005

learning. It can realize the emotion classification of paragraph /section-level learning experience text, so as to mine the implicit emotional states of learners. Kastrati et al.[2] proposes a framework for automatic analysis of opinions expressed by students in comments and automatic recognition of emotions expressed in specific aspects related to MOOC, which greatly reduces the need for manual annotation data.

In addition to text, facial expressions are an important way to show emotions. Jiang et al.[3] used test questions with different difficulties to induce confused emotions in the subjects, and at the same time used camera equipment to capture learners' facial expressions in real time and extract important facial feature points, and then used machine learning algorithm to conduct confusion recognition.

In terms of techniques and methods of sentiment analysis, there are sentiment analysis methods based on representation learning[4], transfer learning[5] and neural network[6]. For example, there are sentiment analysis based on deep representation learning and Gaussian process transfer learning[7], sentiment classification based on representation learning and transfer learning[8], and sentiment analysis method based on convolutional neural network and generative adversarial network[9]. Deep learning has achieved excellent results in various tasks of sentiment analysis[10]. Scholars also merge deep learning and other methods for sentiment analysis. Li et al.[11] use the Bert-CNN model as a review classifier to analyze emotional information in students' course reviews through deep learning, so as to discover problems and improve teaching effects. Dedi et al.[12] combines the uncertainty processing ability of fuzzy logic with the learning ability of deep learning to provide users with more appropriate sentiment prediction. Qing et al.[13] proposed a CNN-LSTM network model to learn the static and dynamic characteristics of group emotions by combining CNN and LSTM for the recognition accuracy of group emotions.

Attention mechanism is the latest achievements in the field of deep learning, which is able to capture the most representative features of the text. A text sentiment analysis model[14] was proposed based on CNN and hierarchical attention network. Attention mechanism[15] was embedded into the convolutional neural network, and a new sentiment analysis method was proposed. Accuracy, stability and efficiency are better than traditional CNN model performance. Lu et al.[16] integrated attention mechanism, BiLSTM and dictionary for sentiment classification and achieved good effect.

3. A CNN-BILSTM WITH ATTENTION MODEL

CNN extracts local features, BiLSTM mines contextual information and an attention mechanism focuses on significant text domains. We integrated the advantages of CNN, BiLSTM and attention mechanism to create a deep learning model. The CNN-BiLSTM model with attention analyzed the online learning comment sentiment.

3.1. Architecture of A CNN-BiLSTM Model with Attention

A CNN-BiLSTM with attention architecture was put forward to analyze sentiment of online learning comments. The architecture is as follows(Figure 1). The architecture has five layers: Embedding layer,CNN layer,BiLSTM layer,attention layer and dense layer.

1)Embedding layer: After data preprocessing, sentences are encoded through indexes, and then embedded matrices are created. Word embedding is a class of approaches for representing words and documents using a dense vector representation.Using embedding matrices can greatly shorten the length of each word vector and setup a framework for our neural network that is able to embed word vector.

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202211_8(11).0005

2)CNN layer: Embedding matrix is fed into CNN layer. Convolution mechanism is applied to extract local features. Pooling mechanism is used to reduce feature dimensions and retain useful information. Here CNN layer include convolution layer and pooling layer.

3)BiLSTM layer: This layer utilizes backward and forward LSTM to model context information and thus enhances the memory capability. It's used for extracting past and future feature. We employ the bi-directional mechanism on learning comments to extract the global feature and obtain information from the context.

4)Attention layer: an attention mechanism for focusing on different words. The attention mechanism learns the importance of each element from the sequence, calculates the probability distribution of the input sequence to the output element to get the weight of allocation, and obtains the importance of the word to the sentence.

5)Dense layer: The last step is flatten the output of attention layer and fed into the dense layer. The dense layer is also called fully connected layer, which is used for realizing the classifying.

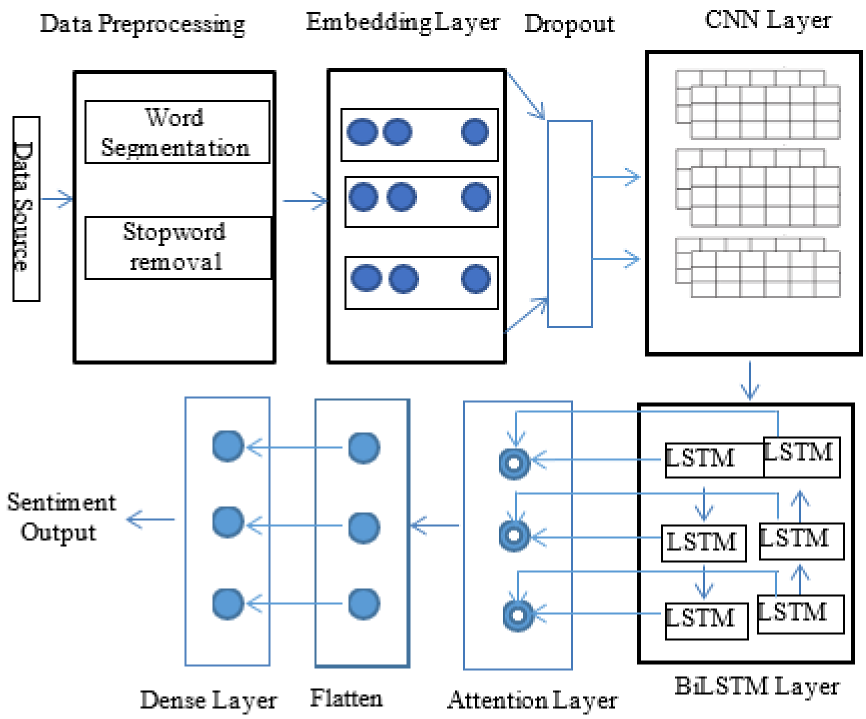

Figure 1. Architecture of CNN-BiLSTM with attention architecture

4. EXPERIMENTS AND PERFORMANCE EVALUATION

4.1. Dataset

We use the web crawler to get the online course comments of MOOC platform and then clean them. Dirty data or junk data were removed, non-standard data were corrected, and emotions were marked. The comments were divided into positive comments and negative comments. We collected 5889 positive comments and 4543 negative comments. A positive comment and a negative comment example are shown in Figure 2 and Figure 3.

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202211_8(11).0005

老师讲得通俗易懂，我觉得老师讲的很好，不拖泥带水，很简洁，但是又讲出了要点与精髓，很棒！赞。
The teacher speaks easy to understand, I think the teacher is very good without dragging, very concise, but also tell the main points and essence, great! Praise.

Figure 2. positive comment example

有些地方讲的比较模糊。
It's kind of vague in some places.

Figure 3. negtive comment example

4.2. Experiment Setup

4.2.1. Experiment steps

Firstly, the data is preprocessed, mainly using JieBa word segmentation tool to segment the Chinese comment text, then removing stop words, making a unified standard for the length of the comment text, completing the parts which are not standard, and truncating the parts that are beyond the standard.

The comment texts labeled I are then expressed as a word vector matrix structure by pre-training word vectors. We use tensorflow built-in preprocessing.VocabularyProcessor to build word to idx mapping relationship. And then use wikipedia pre-training word vector to process vocabulary dictionary to obtain embedding matrix E.

$$E=vocabp(I)$$

After that, the obtained word embedding vectors were dropped out. The purpose of dropout is avoiding overfitting. The output E' of dropout was fed into CNN layer. We get the output M of CNN.

$$M=Conv(E')$$

And the output M of the CNN layer was added into a BiLSTM layer.

$$B=BiLSTM(M)$$

The output B of the Bi-LSTM layer was fed into an attention network.

$$A=Att(B)$$

The last is a fully connected layer. At last we get the sentiment polarity S.

$$S=softmax(A)$$

The CNN-BiLSTM with attention model integrate local features, global features and inter-sentence importance weights, which can obtained a higher level of semantic expression.

We all know that hyper parameters are the most important in the deep learning network. Here the related hyper parameters are as follows.(Table 1.)

Table 1. Hyper parametrs

Hyper parameters	value
batch_size	64
max_sentence_length	200
lstm_output_size	128
filters	250
kernel_size	3
dropout	0.2

4.2.2. Evaluation Criteria

We will do experiments in tensorflow and python language. The Evaluation criteria are r-square(r2), mean absolute error(mae) and loss. The formula of r^2 and mae are formula (1) and (2) in the following. The loss function we use is Keras's 'categorical_Crossentropy' function.

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202211_8(11).0005

$$r^2 = 1 - \frac{\sum_i(\hat{y}_i - y_i)^2}{\sum_i(\bar{y}_i - y_i)^2} \tag{1}$$

$$\mathrm{MAE} = \frac{1}{m}\sum_{i=1}^{m}|(y_i - \hat{y}_i)| \tag{2}$$

In formula (1) and (2) ,y_i is the y true value, $\hat{y}_i$ is the prediction value. $\bar{y}_i$ is the average value of y_i.

Then we will compute the accuracy values with the proposed model in datasets. Meanwhile, we do the experiment and compute the accuracy values with the other model, like naive bayes, SVM, LSTM. And then compare the performance.

4.3. Experiment Result Analysis

4.3.1 Performance Analysis of CNN-BiLSTM with attention

We measure the performance of the model by r^2, MAE, and loss values. The specific visualization results are shown in Figure 4, Figure 5 and Figure 6. The parameter epoch is set 10 .

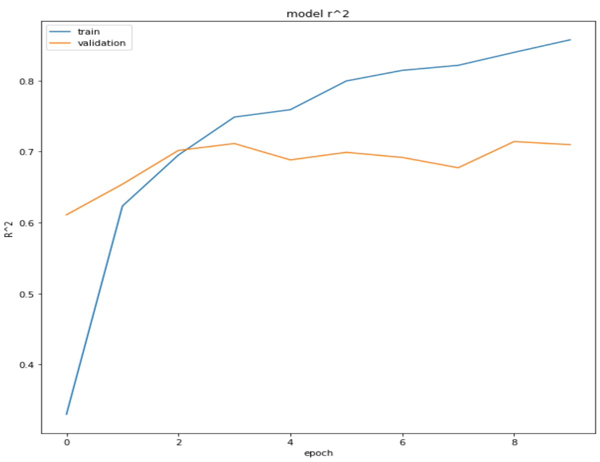

Figure 4. r-square value each epoch

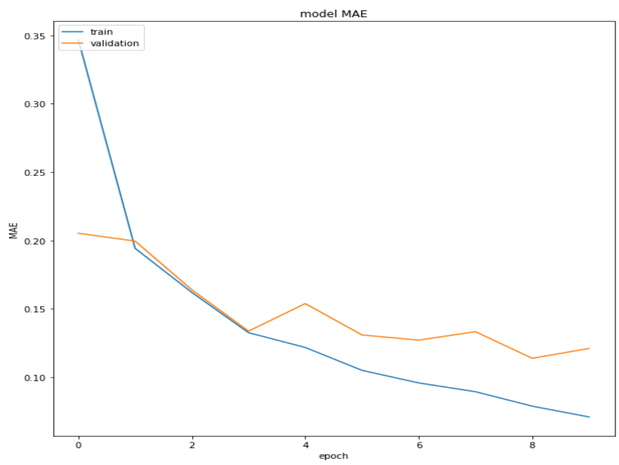

Figure 5. MAE value each epoch

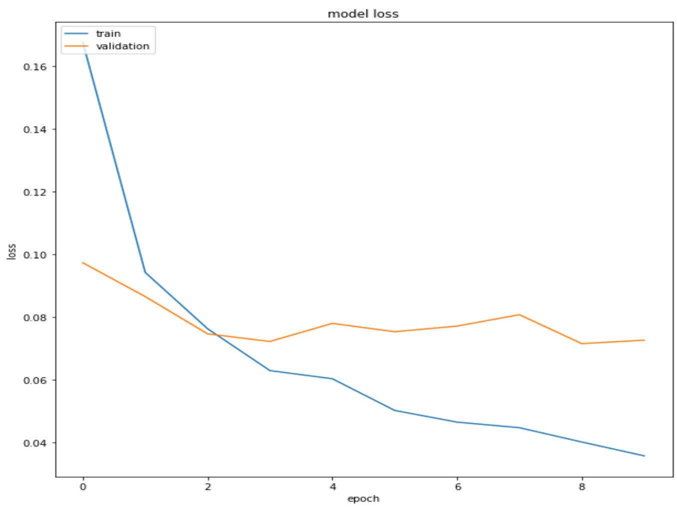

Figure 6. loss value each epoch

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202211_8(11).0005

4.3.2. Accuracy Results and Analysis

We compared our CNN-BilSTM with attention model with naive Bayes model, SVM and LSTM model, and the results are shown in table 1.

It can be seen from the results that the accuracy of our proposed model is 7% higher than that of naive Bayes, 12% higher than SVM and 5% higher than LSTM. The experiments show that our model is better than other models.

Table 2. Accuracy comparison of models

Model	Accuracy
Naive Bayes	0.85
SVM	0.80
LSTM	0.87
CNN-BiLSTM with attention	0.92

5. CONCLUSION AND FUTURE WORK

Our model used a one-dimensional CNN layer on the word embedding layer to extract local features, BiLSTM for extracting long dependencies, and an attention mechanism for focusing on significant text domains. The model integrated benefits of the CNN,BiLSTM and attention and enhanced the performance.

It is of great significance to develop students' learning ability and guide their personalized learning by using online learning comments, mining students' learning state, effectively detecting students' learning emotion and timely adjusting strategies to meet students' learning needs. It is conducive to improving the learning and teaching effect.

We will study the effect of multimodal data on students' emotional polarity in future, including text, audio, video, facial expression, etc. In order to judge students' emotional polarity more accurately, master students' learning state in time, and better serve the teaching work.

ACKNOWLEDGMENTS

This work was supported by Natural Science Research Fund for Universities of Anhui Province (KJ2021A0484) and Philosophy and Social Sciences Fund of Anhui Province (AHSKF2021D31).

REFERENCES

[1] H. Li. Research on Learners' Emotion Analysis Model of Learning Experience Texts,Journal of Distance Education, (2021)No.1,p. 94- 103 (in Chinese)

[2] ZENUN KASTRATI,ALI SHARIQ IMRRAN,ARIANIT KURTI.Weakly Supervised Framework for Aspect-Based Sentiment Analysis on Students' Reviews of MOOCs.IEEE ACCESS, vol.8(2020),p. 106799-106810

[3] B.Jiang,W.J. Li, et al. Automatic Learning Confusion Recognition Using Facial Expressions ,Open education research，vol.24(2018)No.4,p.101-108(in Chinese)

[4] X.J. Li, H.X.Shi, N.N.Chen,et al. Research on sentiment analysis based on representation learning. Acta scientiarum Naturalium Universitatis Pekinensis,vol.55(2019),No.1,p.105-112(in Chinese)

[5] B.Wu ,J. Ji, L.Meng et al. Transfer Learning Based Sentiment Analysis for Poetry of the Tang Dynasty and Song Dynasty, Acta Electronica Sinica, vol.44(2016)No.11,p.2780-2787(in Chinese)

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202211_8(11).0005

[6] Alharbi ASM,de Dondker E.Twitter sentiment analysis with a deep neural network:An enhanced approach using user behavioral information.Cognitive Systems Research,(2019)No.54,p.50-61

[7] D.Y.Wu, L.Gui, Z.Chen, et al. Emotion Analysis Based on Deep Representation Learning and Gaussian Process Transfer Learning，Journal of Chinese Information processing, vol.31(2017) No.1,p.169-176(in Chinese)

[8] X.W.Liao, X.J.Wu,L. GUI, et al. Cross-Domain Sentiment Classification Based on Representation Learning and Transfer Learning, Acta Scientiarum Naturalium Universitatis Pekinensis, vol.55 (2019)No.1,p.37-46(in Chinese)

[9] C.S.Du,L.Huang.Sentiment Analysis Method based on Piecewise Convolutional Neural Network and Generative Adversarial Network.International Journal of Computers Communications & Control, vol.14(2019)No.1,p.7-20.

[10] Y.Lecun,Y.Bengio,G.Hitnton.Deep learning. Nature, vol.521(2015)vol.7553: 436-444.

[11] X. Li,Y.X. Ouyang,W. Rong,et al.A Shallow BERT-CNN Model for Sentiment Analysis on MOOCs Comments. 2019 IEEE International Conference on Engineering (Yogyakarta, Indonesia), Technology and Education,(2019)

[12] P. Dedi,P. Khurana.Sentiment Analysis Using Fuzzy-Deep Learning. Proceedings of ICETIT 2019, Lecture Notes in Electrical Engineering,vol.605(2020),p.246-257.

[13] L.B.Qing,W.S. Xiong,et al. Crowd emotion recognition based on multi-stream CNN-LSTM networks, Application Research of Computer,vol. 35(2018)No.12,p.3828-3831(in Chinese).

[14] Y.Cheng,Z.M. YE,M. W.Wang, et al. Chinese text sentiment orientation analysis based on convolutional neural network and hierarchical attention network. Journal of Chinese Information Processing. Vol.33(2019)No.1,p.133-142(in Chinese).

[15] B.Bonggun Shin,Timothy Lee,Jinho D.Choi.Lexicon Integrated CNN Models with Attention for Sentiment Analysis,Proceeding of the 8th Workshop on Computational Approaches to Subjectivity(Copenhagen,Denmark, September 7-11,2017),Sentiment and Social Media Analysis, p.149-158.

[16] Y.F. Lu,Y.H.Rao,J.Yang, et al. Incorporating Lexicons into LSTM for Sentiment Classification, Proceedings of the International Joint Conference on Neural Networks(Rio de Janeiro, Brazil (2018.7.8-2018.7.13).(2018),p.1-7.