

Article

# Sentiment Analysis of Comment Texts on Online Courses Based on Hierarchical Attention Mechanism

Baohua Su <sup>1,2</sup> and Jun Peng <sup>2,\*</sup>

<sup>1</sup> College of Chinese Language and Culture, Jinan University, Guangzhou 510632, China; subaohua@hwy.jnu.edu.cn

<sup>2</sup> School of Education, Research Institute of Macau Education Development, City University of Macau, Macau 999078, China

\* Correspondence: 4588775@163.com

**Abstract:** With information technology pushing the development of intelligent teaching environments, the online teaching platform emerges timely around the globe, and how to accurately evaluate the effect of the “any-time and anywhere” teacher–student interaction and learning has become one of the hotspots of today’s education research. Bullet chatting in online courses is one of the most important ways of interaction between teachers and students. The feedback from the students can help teachers improve their teaching methods, adjust teaching content, and schedule in time so as to improve the quality of their teaching. How to automatically identify the sentiment polarity in the comment text through deep machine learning has also become a key issue to be automatically processed in online course teaching. The traditional single-layer attention mechanism only enhances certain sentimentally intense words, so we proposed a sentiment analysis method based on a hierarchical attention mechanism that we called HAN. Firstly, we use CNN and LSTM to extract local and global information, gate mechanisms are used for extracting sentiment words, and the hierarchical attention mechanism is then used to weigh the different sentiment features, with the original information added to the attention mechanism concentration to prevent the loss of information. Experiments are conducted on China Universities MOOC and Tencent Classroom comment data sets; both accuracy and F1 are improved compared to the baseline, and the validity of the model is verified.

**Keywords:** review text for online courses; sentiment analysis; attention mechanism; gating mechanism



**Citation:** Su, B.; Peng, J. Sentiment Analysis of Comment Texts on Online Courses Based on Hierarchical Attention Mechanism. *Appl. Sci.* **2023**, *13*, 4204. <https://doi.org/10.3390/app13074204>

Academic Editor: João M. F. Rodrigues

Received: 2 March 2023

Revised: 23 March 2023

Accepted: 24 March 2023

Published: 26 March 2023



**Copyright:** © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

In recent years, network technologies, such as the Internet, the Internet of things, and big data, have developed rapidly, and network platforms for e-commerce, social communication, and education are emerging timely. These platforms have not only enriched our daily life but also changed our ways of working, studying, and living. The sentiment comment texts on the network platform reflect people’s opinions on something. Thus, how to effectively use these opinions has become an important factor in improving service quality. In education, many countries have shifted their offline teaching to online teaching due to the global COVID-19 pandemic [1,2]. Compared with the traditional offline classroom, online education has the advantages of lower costs, flexible forms, and fewer geographical restrictions [3,4]. Its promotion and application increase the equity of higher education, realize knowledge sharing, improve the effectiveness and efficiency of decision-making, and make higher education more open [5]. In order to further evaluate the quality of teaching and strengthen the interaction between teachers and students, a large number of teaching platforms, such as China Universities MOOC and Tencent Classroom, have provided the bullet chatting function. The bullet chatting imbued with sentiment information plays an important role in the teaching process. Through students’ feedback, teachers can know what points students are weak in. School administrators can dynamically adjust the knowledge points, teaching plans, teaching objectives, and teaching staff structure of the courses

based on the sentiment analysis of comment texts. Therefore, how to leverage useful information from comment text with sentiment information has become one of the hot research directions in natural language processing [6].

Sentiment analysis is used to judge the sentiment polarity (positive, neutral, or negative) of reviews. Since Pang et al. studied the sentiment analysis of film reviews, sentiment analysis technology has been widely used in the business community [7]. As an emerging educational approach in the era of information technology, online courses have attracted many educators and learners around the world with their advantages of spanning time and space and flexible learning methods. Comments, as the most direct way of interactive feedback in online courses, are of great significance in improving the quality of teaching, reducing the dropout rate, and promoting the sustainable development of online courses [8–10]. So, sentiment analysis is also very important in the field of education, but very few researchers do sentiment analysis in online course reviews, and even public data sets on this are very scarce.

There are three main methods of sentiment analysis: sentiment analysis based on sentiment dictionaries and rules, sentiment analysis based on traditional machine learning, and sentiment analysis based on deep learning [11]. Soe et al. further calculated sentiment scores to achieve the purpose of analyzing students' emotions through a part-of-speech tagging analyzer and vocabulary resources [12]. The second type of sentiment analysis method recognizes sentiment through constructing features artificially and using naïve Bayes, maximum entropy, and support vector machine and other classifiers. The accuracy of this method depends entirely on the construction of features and the selection of classifiers, while most of the current research mainly focuses on the former. Therefore, the quality of feature selection largely determines the accuracy of the experimental results. Feature construction faces not only the problems of large workload and sparsity of features but also the common problem of domain adaptability.

With the development of deep learning and the improvement in text representation methods based on deep learning, many researchers began to study the application of deep learning to deal with text sentiment analysis. Represented by RNN, LSTM, and other classical neural networks, deep learning-based sentiment analysis methods can not only solve the shortcomings of traditional machine learning but also have significant classification effects. CNN can obtain the local information of a text, whereas recurrent neural networks such as LSTM can obtain the global information of a text. On the one hand, sequence-based neural networks such as LSTM have been restricted by the sequence length and computational memory. Attention mechanisms, on the other hand, could alleviate this problem since it allows modeling of the dependency output sequence without considering the distance between texts [13–15]. As a result, there are some sentiment analysis methods that use classical neural networks combined with an attention mechanism. Yang et al. [16] and Liu et al. [17] shows that the combination of an attention mechanism and LSTM can improve the accuracy of the model. The single-layer attention mechanism tends to focus on the words with strong sentiment expression and ignore the words with weak sentiment expression and opposite polarity, leading to the misjudgment of sentiment polarity. Take a real comment, for example:

“互联网时代,教师个人知识与在线资源连线,现在的问题不是资源太少,而是资源太多,良莠混杂,无从选择。该课程讲解了教师个人知识管理的体系架构,资源分类与统筹管理的方法,对教师理清个人知识体系,提高工作学习效率大有裨益。(In the Internet era, teachers' personal knowledge is connected with online resources. The problem now is not that there are too few resources, but that there are too many, and it is hard to choose from these resources as they are of mixed qualities. This course explains the system structure of teachers' personal knowledge management, the classification of resources and integrated approaches to management, which is of great benefit for teachers to clarify their personal knowledge system and improve their work and learning efficiency.)”

As the single-layer attention mechanism often focuses on the words with strong sentiment expressions, such as “资源太少 (too few resources),” “良莠混杂 (mixed resources differ in quality),” and “无从选择 (there is no way to choose),” resulting in the misjudgment of sentiment polarity. Therefore, this paper proposes the use of a hierarchical attention mechanism to deal with the sentiment analysis of online course reviews. Though CNN and LSTM are often combined for text sentiment analysis, such as [18,19], there is still a lack of effective ways to take advantage of useful information (e.g., historical, global, and local information) extracted by them. Therefore, this paper uses the gate mechanism to further select useful local information. The significance of this study may be declared as follows:

First, this study indicated that the single-layer attention mechanism cannot accurately identify the sentiment words that are useful for global information. When human beings know that they need to carry out sentiment analysis task of short texts, they will first pay attention to the sentiment words in the sentence and then read the sentence from the beginning to the end to judge which sentiment words are more important, and to obtain the sentiment polarity of the sentence. According to this and the human’s way of sentiment analysis of reading text, this study designs a hierarchical interactive attention mechanism, obtains the local features of sentences through CNN, and obtains the global information and the temporal features of sentences by using LSTM. Then the gate mechanism filters the local sentiment information. At the same time, the local sentiment word information extracted by CNN can enrich the hidden layer representation extracted by LSTM, and then the sentiment polarity of the sentence can be obtained after the information is weighted by the hierarchical attention mechanism.

Second, in the design of the attention mechanism, this study preserves original information through the connection way of residuals. The experiment shows that the hierarchical attention mechanism is effective in the sentiment analysis of online course reviews.

## 2. Related Work

Based on deep learning, there are two major categories of sentiment analysis models: graph-based models and sequence-based models.

The TextGCN model proposed by Yao et al. was the first time to use GCN in text classification (sentiment analysis) [20]. Two graphs were employed by that study as effective tools. The one named PMI was used to construct the relationship between words, and another named TF-IDF was used to construct the relationship between documents and words, and then the text category was obtained by the classifier. Then, Ragesh et al. [21] and Galke et al. [22] developed HeteGCN, which combined features of predictive text and TextGCN; It means the adjacency matrix was split into word documents and word submatrices, and the representations of different layers were fused as needed. Subsequently, HyperGAT was brought forward by Ding et al., from which an edge can connect multiple vertices [23]. So the text information was transformed into a hypergraph between nodes and edges, and the information between each layer was aggregated by dual attention. At last, tensorGCN was presented by Liu et al. [24]. This model constructed multiple graphs to describe semantic, syntactic, and contextual information and improved the effect of text classification through learning intra-graph propagation and inter-graph propagation.

Some studies have found that in recent years, most of the new methods for sentiment analysis (text classification) are based on GCN, while transformer-based sequence models are rare in the literature [22]. However, much empirical evidence shows that transformer-based sequence models outperform GCN-based methods. So here is a look at some sequence-based text classification methods. After obtaining the representation of each word, Kim embedded the word into CNN to obtain the sentiment polarity of the text [25]. Through the experimental results of a large number of data sets, he proved the ability of CNN on the task of text classification. After obtaining the text representation, Liu et al. used an RNN to classify the sentiment of the comment text [26]. Wang et al. proved that LSTM could achieve better experimental results than traditional RNNs in

tweet sentiment analysis through experiments on tweet datasets [27]. After acquiring the word representation, the RNNs acquire the phrase representation and the sentence representation in order according to the syntactic structure. Huang et al. used a two-layer LSTM to classify the sentiment of tweets and believed that the sentiment polarity of the current tweet was largely related to the previous and subsequent tweets [28]. If the sentiment polarity is judged by the current tweet alone, the system would be deceived by its irony and other language expressions. Therefore, the hidden layer state of the current tweet should be input into a higher-level LSTM to obtain the current tweet representation containing context information and, finally, obtain the sentiment polarity distribution of the current tweet through the classifier. Yang et al. used the attention mechanism to aggregate word information to obtain sentence information, then they used the second layer attention mechanism to aggregate sentence information, in order to obtain the overall sentiment polarity in the discourse-level sentiment analysis, which fully proved the importance of an attention mechanism in sentiment analysis [16]. Vaswani et al. proposed the transformer model, which once again proved the importance of an attention mechanism in text classification [13]. Since the invention of BERT in 2018, there has been a lot of research on sentiment analysis based on BERT [29]. In Order to solve the negative effect of mask in BERT, XLNet uses an autoregressive language model instead of an autoencoding language model and introduces a double-stream self-attention mechanism and transformer-xl [30]. Compared with BERT, XLNet achieves better experimental results. ERNIE uses the same coding structure as BERT, but the author thinks that the random mask mechanism in BERT ignores the semantic relationship to some extent, so the original mask is split into three parts, the first part retains the original random mask, the second part masks the entity word as a whole. The last part is to mask the phrase as a whole. Compared with ERNIE, ERNIE 2.0 proposes three types of unsupervised tasks, which provide the model with a better representation ability of sentences, grammar, and semantics [31]. The performance and advantages of some methods on data sets are summarized in Table 1.

**Table 1.** Comparison with some methods.

Model	Data (acc)						Advantages
	SST-2	20NG	R8	R52	Ohsumed	MR	
Text GCN	-	0.863	0.970	0.935	0.683	0.767	A heterogeneous graph based on text and words is constructed, and the semi-supervised classification of text can be performed on GCN
HeteGCN	-	0.846	0.972	0.939	0.638	0.756	Reduce the complexity of TextGCN
HyperGAT	-	0.862	0.970	0.950	0.699	0.783	Capturing higher-order interactions between words while improving computational efficiency
TensorGCN	-	0.877	0.980	0.951	0.701	0.780	Rich multi-subgraph feature representation
LSTM	-	0.754	0.961	0.905	0.511	0.773	More effective way to process sequence data
BERT	0.928	-	-	-	-	-	The vector representation is rich, which overcomes the gradient problem of LSTM when solving long sequence data
ROBERTa	0.937	-	-	-	-	-	Raining models with larger corpora and sequences, dynamic MASK mechanism
XL-net	0.971	-	-	-	-	-	Autoregressive training method to overcome the shortcomings of bert
ernie	0.935	-	-	-	-	-	Taking advantages of The lexical, syntactic and knowledge information, large-scale text corpora and KGs to train an augmented language representation model

“-” indicates that the original paper was not tested on this data set.

### 3. Model Building

A segment with a length of  $n$  is given when we analyze the sentiment of online course reviews. Additionally, the sentiment polarity expressed by the bullet screen is judged by

analyzing the review. After obtaining the comment sentence  $S$ , each word in the sentence needs to be vectorized. In this study, each word in the sentence  $S$  is randomly initialized. That is  $S_v = [v^1, v^2, \dots, v^n] \in \mathbb{R}^{n \times d_w}$ , in which  $d_w$  is the size of the word vector dimension.

### 3.1. Model Construction Process

In order to obtain the local sentiment features of reviews, we use CNN to obtain the local features of sentences (i.e.,  $H_C = [h_c^1, h_c^2, \dots, h_c^m] \in \mathbb{R}^{m \times d_c}$ ) as shown in Figure 1. Then, to further extract the hidden features of the text, LSTM is used to extract the hidden information of the comment text. Finally, the context hidden state (i.e.,  $H_L = [h_l^1, h_l^2, \dots, h_l^n] \in \mathbb{R}^{n \times d_h}$ ) is extracted by LSTM. After that, a gate mechanism is used to screen the important sentiment information of  $H_C$  and  $H_L$ , as follows.

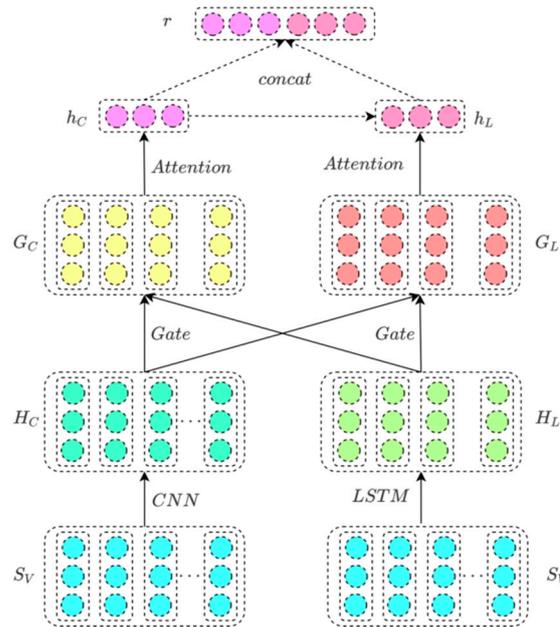


Figure 1. Overall model structure.

First, average pooling needs to be performed for the hidden state of the sentence  $H_L$  extracted by LSTM and the local emotion information  $H_C$ , as shown in Formula (1) and Formula (2), respectively. The LSTM hidden layer vector representation  $h_g^L \in \mathbb{R}^{d_h}$  is obtained. Additionally, a local aggregate information vector representation  $h_g^C \in \mathbb{R}^{d_c}$  is obtained.

$$h_g^L = \sum_{i=1}^n H_L / n \tag{1}$$

$$h_g^C = \sum_{i=1}^m H_C / m \tag{2}$$

Then the local emotion information extracted from CNN needs to be filtered through the gate mechanism by using  $h_g^L$ , and the specific gate mechanism calculations are shown in Formulas (3)–(5).

$$T_C = \text{relu}(H_C W_C + W_g h_g^L \times w_g) \tag{3}$$

$$E^i = \text{tanh}(H_C^i W_E) \tag{4}$$

$$G_C^i = E^i T_C \tag{5}$$

In these calculations,  $W_C \in \mathbb{R}^{d_c \times d_c}$ ,  $W_g \in \mathbb{R}^{m \times d_h}$ ,  $w_g \in \mathbb{R}^{d_c}$  are parameter vectors;  $\text{relu}$ ,  $\text{tanh}$  are activation functions; and  $W_E \in \mathbb{R}^{d_c \times m}$ ,  $G_C^i \in \mathbb{R}^{d_c}$  are global features. A selec-

tive representation of the gate mechanism,  $G_C \in \mathbb{R}^{m \times d_c}$ , can be obtained after expressing the chosen  $i$  vector of  $H_C$  through the gate mechanism.

$G_L \in \mathbb{R}^{n \times d_h}$  can be obtained in the same way. After  $G_C$  is obtained, information is aggregated by the attention mechanism, as shown in Formulas (6) and (7).

$$\alpha = softmax (G_C w_\alpha + H_C w_C) \tag{6}$$

$$h_c = \sum_{i=1}^m \alpha^i G_C^i \tag{7}$$

$w_\alpha \in \mathbb{R}^{d_c}$ ,  $w_C \in \mathbb{R}^{d_c}$  are parameter vectors, and  $H_C$  is the feature information extracted by the original CNN. After obtaining the selection information  $G$  of the gate mechanism, the original information  $H_C$  is further added when the attention mechanism coefficients are weighted to avoid the loss of original information. Finally, the vector  $h_C \in \mathbb{R}^{d_h}$  is weighted by Formula (7).

The single-layer attention mechanism can only focus on the strong sentiment words in the sentiment expression while ignoring the most important words in sentiment analysis. In order to highlight the importance of different words in sentiment analysis, this study proposes to use a multi-layer attention mechanism to focus on the importance of different words in sentences; therefore, weighted by the first layer of attention mechanism  $h_c$ . After that, to improve the accuracy of sentiment analysis by using text information, the information of  $G_L$  is further weighted by  $h_c$ , as shown in Formulas (8)–(10).

$$\gamma (h_L^i, h_c) = tanh (h_L^i W_L h_c^T) \tag{8}$$

$$\beta_i = \frac{exp (\gamma (h_L^i, h_c))}{\sum_{j=1}^n exp (\gamma (h_L^j, h_c))} \tag{9}$$

$$h_t = \sum_{i=1}^n \beta_i h_L^i \tag{10}$$

$h_L^i \in \mathbb{R}^{d_h}$  is the  $i$  vector of  $G_L$ ,  $tanh$  is the activation function,  $W_L \in \mathbb{R}^{d_h \times d_c}$  is parameter matrix, and  $h_c^T$  is the transpose of the  $h_c$  vector.  $\gamma (h_L^i, h_c)$  is obtained by Formula (8), and  $\gamma (h_L^i, h_c)$  is the attention mechanism coefficient between  $h_L^i$  and  $h_c$ .  $\beta_i$  is the attention mechanism coefficient between  $h_L^i$  and  $h_c$  after they are normalized. Finally,  $h_t \in \mathbb{R}^{d_h}$  is obtained by the weighted summation of Formula (10).

$r \in \mathbb{R}^{d_c+d_h}$  can be obtained by splicing  $h_c$  and  $h_t$ , and it is the final representation of a sentence containing sentiment information. Finally, the sentiment polarity of the sentence is obtained by the *softmax* classifier, as shown in Formulas (11) and (12).

$$x = tanh (W_r r + b_r) \tag{11}$$

$$y_i = \frac{exp (x_i)}{\sum_{j=1}^C exp (x_j)} \tag{12}$$

$W_r \in \mathbb{R}^{C \times (d_c+d_h)}$  is a parameter matrix,  $b_r \in \mathbb{R}^C$  is a bias vector, and  $C$  is the total number of sentiment categories. Two data sets are used in this study, and the total number of sentiment classifications of one data set is two, positive and negative. The total number of sentiment classifications in the other data set is three, and they are positive, neutral, and negative.

### 3.2. Model Training

In order to use backward propagation to iteratively update all the parameter matrices and bias vectors proposed above, this study uses the cross-entropy and regularization of all

the sentence classification results of the training set as the loss function, and the formulas are shown in (13) and (14).

$$J = - \sum_{i=1}^C g_i \log y_i + \lambda_r \left( \sum_{\theta \in \Theta} \theta^2 \right) \quad (13)$$

$$\Theta = \Theta - \lambda_l \frac{\partial J(\Theta)}{\partial \Theta} \quad (14)$$

$g_i$  is the true sentiment distribution in the reviews,  $y_i$  is the prediction of the sentiment polarity of the review by the model,  $\Theta$  is the set of all parameters,  $\lambda_r$  is the parameter of the  $L_2$  regularization, and  $\lambda_l$  is the learning rate of the update parameter.

#### 4. Experimental Process

In order to verify the effect of the proposed multi-layer attention mechanism of sentiment analysis model, this study conducted experiments, including data sets, evaluation indicators, and hyperparameter settings.

The average accuracy and F1 are calculated as shown in Formulas (15)–(18).

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \quad (15)$$

$$precision = \frac{TP}{TP + FP} \quad (16)$$

$$recall = \frac{TP}{TP + FN} \quad (17)$$

$$F1 = \frac{2 \times precision \times recall}{precision + recall} \quad (18)$$

where  $TP$  means that the real label is a positive example and the predicted label is also a positive example;  $TN$  means that the real label is a negative example and the predicted label is also a negative example;  $FP$  means that the true label is a negative example, and the predicted labels are positive examples;  $FN$  means that real labels are positive examples and predicted labels are negative examples. The average accuracy represents the proportion of the correct prediction to all data, precision represents the accuracy of the positive prediction, recall represents the proportion of the correct prediction to all positive examples, and F1 takes into account both accuracy and recall, which is also a commonly used measurement standard.

##### 4.1. Experimental Setup

**Data sets:** The data sets used in this study are two open online course review data sets, namely China University MOOC reviews [32] and MOOC and Ke reviews [25]. China University MOOC is jointly launched by NetEase Youdao and Higher Education Press and carries more than 10,000 open courses and more than 1400 national quality courses. It cooperates with 803 universities and is the largest Chinese MOOC platform [26]. Ke reviews comes from Tencent Classroom. Tencent Classroom is a comprehensive online lifelong learning platform launched by Tencent. It gathers a large number of high-quality educational institutions and famous teachers and offers many online quality courses [25]. Such as vocational training, civil service examination, TOEFL and IELTS, certification and grading examination, oral English, etc. Both MOOC and Tencent Classroom in Chinese universities contain a large number of users and rich classroom reviews. Sentiment analysis of classroom comments can master students' sentiment tendencies and help to carry out targeted classroom improvement, which can improve teaching quality to a certain extent. The MOOC data set contains 11,293 reviews on Chinese online courses from MOOCs, and the affective polarity of the data is divided into positive and negative, with 6164 positive

reviews and 5129 negative reviews. MOOC and Ke reviews collected 1808 online course reviews from MOOC and Ke reviews, and the affective polarity of the data set was classified as positive, neutral, and negative. Among them, there are 817 reviews on positive sentiment polarity, 750 reviews on neutral sentiment polarity, and 241 reviews on negative sentiment polarity. In this study, the two data sets were mixed. For the MOOC data set, 80% of them were set as the training set, 10% as the validation set, and 10% as the test set. Due to the small number of MOOC and KE data sets, 80% of them are set as the training set, and the rest are set as the test set. The specific distribution of the two data sets is shown in Table 2. In addition, experiments on the common data set R8 were conducted. The total number of data in R8 is 7674, among which the number of data in the training set is 5482, the number of data in the test set is 2189, and the classification category is 8.

**Table 2.** Data set distribution.

Dataset	Positive	Neutral	Negative
MOOC-Train	4609	0	4068
MOOC-Val	598	0	531
MOOC-Test	600	0	630
MOOC and Ke-Train	639	609	200
MOOC and Ke-Test	180	141	41

Evaluation index: Average accuracy is used to measure the performance of the sentiment analysis model of online classroom reviews based on a hierarchical attention mechanism, and F1 is used to evaluate it on MOOC.

Hyperparameters: We use a single Chinese character as a word. Words are initialized through random vector initialization; the dimension of the word vector is 300, the dimension of the LSTM hidden layer is also set to 300, and the number of LSTM layers is set to 2. The convolution windows of CNN are 2, 3, and 4, and the number of convolution sums is set to 256. For the words not in the dictionary, the values are randomly taken between  $-0.1$  and  $0.1$ . The initial values of all the parameter matrices and vectors are randomly chosen between  $-0.1$  and  $0.1$ . The initial value of the bias is set to 0. To adjust the parameters, the optimizer used in this study was Adam, the learning rate was set to 0.01, and the dropout was set to 0.5 to prevent overfitting.

#### 4.2. Experimental Results and Analysis

The results of the comparison between the seven baselines and the HAG are shown in the following Table 3.

**Table 3.** Comparative experimental results.

Methods	MOOC (acc and F1)	MOOC and Ke (acc)	R8
CNN	0.903 and 0.911	0.453	95.34
LSTM	0.912 and 0.918	0.461	96.09
LSTM-Attention	0.932 and 0.920	0.472	96.59
BERT	0.932 and 0.940	0.495	98.03
RoBERTa	0.934 and 0.945	0.496	98.23
ERNIE	0.937 and 0.936	0.496	98.04
HAN	0.940 and 0.938	0.499	97.65

It can be seen from Table 2 that the experimental results of all models on the MOOC and MOOC and Ke data sets are quite different for four reasons: First, the data of MOOC is relatively large compared with the data of MOOC and Ke, and the number of data in MOOC is 11,036, while the number of data in MOOC and Ke is 1810. Secondly, the MOOC and Ke data set is divided into positive and negative categories, while the reviews in MOOC and Ke are divided into positive, neutral, and negative categories, which increases the difficulty of classification to a certain extent. Then, the proportion of positive

and negative MOOCs relative to the MOOC and Ke dataset is more balanced. Finally, the review data in MOOC is clean, while the MOOC and Ke data set is noisy, which limits the accuracy of sentiment classification of MOOC and Ke data to some extent. In these three data sets, the accuracy of CNN is lower than in other models. CNN plays a significant role in extracting local features and can obtain sentiment word information to a certain extent. LSTM can learn long-term dependencies and extract the sequence information of text effectively. Compared with CNN, the accuracy of LSTM on MOOC, MOOC, Ke, and R8 are increased by 0.9%, 0.8%, and 0.75%, respectively, and the accuracy of F1 is increased by 0.07%. The accuracy of the model is improved by the attention mechanism. BERT is also a classic model in sentiment analysis. BERT uses a multi-head attention mechanism to give the output of the attention layer, which contains the coding representation information in different subspaces, thus enhancing the expressive power of the model. Since the use of BERT for emotion analysis, various improvement methods based on BERT have been proposed. RoBERTa uses a dynamic masking mechanism and abandons that NSP (Next Sentence Predict) task; compared with BERT, RoBERTa performs slightly better on both datasets. Ernie also adjusts the mask mechanism in BERT. In the sentiment analysis of Chinese online courses, Ernie can identify the importance of words better than BERT. HAN is inferior to BERT-based models (i.e., BERT, Roberta, and Ernie) on long news texts of R8. This is partly due to the randomly initialized representation of HAN. Moreover, for the long news texts, BERT could alleviate the problem of vanishing gradients. However, in the class comments, HAN has achieved the best experimental results. HAN first uses CNN to extract local sentiment information and then uses the gate mechanism to filter the local sentiment information through the overall text information obtained by LSTM. At the same time, it uses the local sentiment information extracted by CNN to enrich the sequence information extracted by LSTM. Then through the weighting of the hierarchical attention mechanism, the sentiment tendency experiment of online classroom network reviews is obtained, and the experimental results once again prove the effectiveness of the HAN proposed in this study.

### 4.3. Case Study

To test the reliability of the model, a visual comparison between the weight of the final attention mechanism and the weight of the attention mechanism in baseline LSTM-Attention was made. The sample is derived from the comments in real online courses, such as Figure 2. As shown, the top half is the weight coefficient of the proposed model, and the bottom half is the weight of the attention mechanism of LSTM-Attention. The darker color means the greater weight of the attention mechanism and vice versa.

“她们的许许多多的创意都很值得我学习, 但会自责, 同样的课程, 为什么差距这么大。(Most of their creative ideas are well worth learning, but we will reproach ourselves for the same course with a great difference.)”

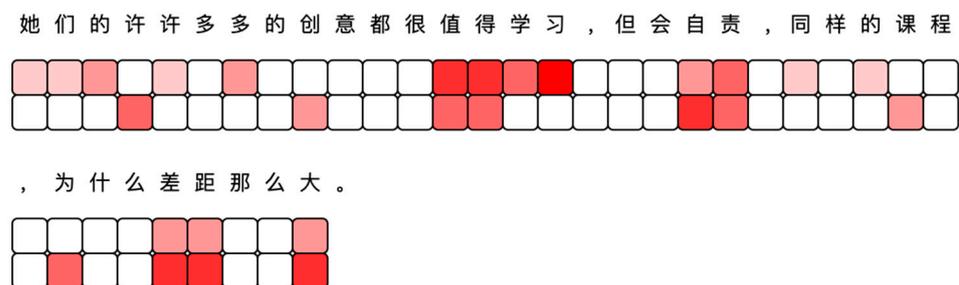


Figure 2. Example for the comments of real online course.

It seems that in Figure 2, the words “值得学习 (worth learning)”, “自责 (reproach ourselves)”, and “差距 (difference)” possess larger attention mechanism weights. However, when taking a closer look, it can be noticed that the overall emotional tendency of the sam-

ple comment is positive. Meanwhile, negative words such as “自责” and “差距大 (a great difference)” would interfere with the result. In this case, The LSTM-Attention method cannot distinguish the effect of these words on the overall sentiment analysis, and the words with strong emotional intensity were given higher weights. The model proposed in this study strengthened the weight of “值得学习” and correspondingly reduced the weight of “自责”, “大”, and other words, which verified the effectiveness of the model.

## 5. Conclusions

Using deep learning technology and starting from the sentiment analysis of online course review text, this study proposes a method to analyze the comments on online courses based on a hierarchical attention mechanism. The method enriches the extracted information by using CNN to extract local sentiment information and LSTM to obtain the hidden representation of the text. Then the global sentiment information extracted by the CNN and the global information extracted by the LSTM are screened by gate mechanism, respectively. The hierarchical attention mechanism reduces the influence of noise on affective polarity judgments, and the interference of the words with strong sentiment information to the model judgment is also reduced. This study proves the reliability of the HAN through three data sets. In the future, we will use more information to enrich the embedded representation of words, such as adding some speech information, sentiment information, and location information. In addition, in the construction of data sets, we will collect high-quality attribute-level online course reviews to ensure the accuracy of online courses' information feedback and promote the courses' deep interaction.

**Author Contributions:** Conceptualization, B.S. and J.P.; methodology, B.S.; software, J.P.; validation, B.S. and J.P.; formal analysis, B.S.; investigation, B.S.; resources, B.S. and J.P.; data curation, B.S. and J.P.; writing—original draft preparation, B.S.; writing—review and editing, B.S. and J.P.; visualization, B.S.; supervision, J.P.; project administration, B.S. and J.P.; funding acquisition, J.P. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was financially supported by the 2021 Higher Education Fund of the Macao SAR Government. (Project name: Development and Effectiveness Assessment of Higher Education Online Courses in Macao. Project No: HSS-CITYU-2021-07).

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The data used to support the findings of this study can be obtained from the corresponding author upon request.

**Conflicts of Interest:** The authors declare that there are no conflict of interest regarding the publication of this paper.

## References

1. Tarkar, P. Impact of COVID-19 pandemic on education system. *Int. J. Adv. Sci. Technol.* **2020**, *29*, 3812–3814.
2. Zhou, L.; Wu, S.; Zhou, M.; Li, F. 'School's out, but class' on', the largest online education in the world today: Taking China's practical exploration during The COVID-19 epidemic prevention and control as an example. *Best Evid. Chin. Edu.* **2020**, *4*, 501–519. [[CrossRef](#)]
3. Cao, R.; Xu, S.; Wang, X. Digitalization Leads the Future of Global Higher Education—Summary of the Main Session of the 2022 World MOOC and Online Education Conference. *China Educ. Informatiz.* **2023**, *29*, 82–95. [[CrossRef](#)]
4. Wang, X.; Guo, S. Practice and Enlightenment of Online and Offline Integrated Teaching in Tsinghua University. *Mod. Educ. Technol.* **2022**, *32*, 106–112.
5. Global MOOC and Online Education Alliance. Trends, Stages and Changes of Digitalization of Higher Education: An Excerpt from Infinite Possibilities: Report on the Development of Digitalization of World Higher Education. *China Educ. Informatiz.* **2023**, *29*, 3–8. [[CrossRef](#)]
6. Feng, C.; Li, H.; Zhao, H.; Xue, Y.; Tang, J. Attribute level sentiment analysis based on hierarchical attention mechanism and gate mechanism. *Chin. J. Inf. Technol.* **2021**, *35*, 128–136.
7. Pang, B.; Lee, L.; Vaithyanathan, S. Thumbs up? Sentiment classification using machine learning techniques. *arXiv* **2002**, arXiv:cs/0205070.

8. Wang, L.; Hu, G.; Zhou, T. Semantic analysis of learners' sentiment tendencies on online MOOC education. *Sustainability* **2018**, *10*, 1921. [[CrossRef](#)]
9. Mite-Baidal, K.; Delgado-Vera, C.; Solís-Avilés, E.; Espinoza, A.H.; Ortiz-Zambrano, J.; Varela-Tapia, E. Sentiment analysis in education domain: A systematic literature review. In Proceedings of the Technologies and Innovation: 4th International Conference, CITI 2018, Guayaquil, Ecuador, 6–9 November 2018; Springer International Publishing: Cham, Switzerland, 2018; pp. 285–297.
10. Pan, F.; Zhang, H.; Dong, J.; Shou, Z. Sentiment analysis of Chinese online course reviews based on efficient Transformer. *Comput. Sci.* **2021**, *48*, 264–269.
11. Xu, G.; Meng, Y.; Qiu, X.; Yu, Z.; Wu, X. Sentiment analysis of comment texts based on BiLSTM. *IEEE Access* **2019**, *7*, 51522–51532. [[CrossRef](#)]
12. Soe, N.; Soe, P.T. Domain oriented aspect detection for student feedback system. In Proceedings of the 2019 International Conference on Advanced Information Technologies (ICAIT), Yangon, Myanmar, 6–7 November 2019; pp. 90–95.
13. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, Ł.; Polosukhin, I. Attention is all you need. In Proceedings of the 31st International Conference on Neural Information Processing Systems, Long Beach, CA, USA, 4–9 December 2017; Volume 30.
14. Bahdanau, D.; Cho, K.; Bengio, Y. Neural machine translation by jointly learning to align and translate. *arXiv* **2014**, arXiv:1409.0473.
15. Kim, Y.; Denton, C.; Hoang, L.; Rush, A.M. Structured attention networks. In Proceedings of the International Conference on Learning Representations, Toulon, France, 24–26 April 2017.
16. Yang, Z.; Yang, D.; Dyer, C.; He, X.; Smola, A.; Hovy, E. Hierarchical attention networks for document classification. In Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, San Diego, CA, USA, 12–17 June 2016; pp. 1480–1489.
17. Liu, Z.; Zhou, W.; Li, H. AB-LSTM: Attention-based bidirectional LSTM model for scene text detection. *ACM Trans. Multimed. Comput. Commun. Appl. (TOMM)* **2019**, *15*, 107. [[CrossRef](#)]
18. Zhang, J.; Li, Y.; Tian, J.; Li, T. LSTM-CNN Hybrid Model for Text Classification. In Proceedings of the 2018 IEEE 3rd Advanced Information Technology, Electronic and Automation Control Conference (IAEAC), Chongqing, China, 12–14 October 2018; pp. 1675–1680. [[CrossRef](#)]
19. She, X.; Zhang, D. Text classification based on hybrid CNN-LSTM hybrid model. In Proceedings of the 2018 11th International Symposium on Computational Intelligence and Design (ISCID), Hangzhou, China, 8–9 December 2018; Volume 2, pp. 185–189.
20. Yao, L.; Mao, C.; Luo, Y. Graph convolutional networks for text classification. *Proc. AAAI Conf. Artif. Intell.* **2019**, *33*, 7370–7377. [[CrossRef](#)]
21. Ragesh, R.; Sellamanickam, S.; Iyer, A.; Bairi, R.; Lingam, V. Hetegcn: Heterogeneous graph convolutional networks for text classification. In Proceedings of the 14th ACM International Conference on Web Search and Data Mining, Virtual Event, Israel, 8–12 March 2021; pp. 860–868.
22. Galke, L.; Scherpp, A. Bag-of-words vs. graph vs. sequence in text classification: Questioning the necessity of text-graphs and the surprising strength of a wide MLP. In Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics, Dublin, Ireland, 22–27 May 2022; pp. 4038–4051.
23. Ding, K.; Wang, J.; Li, J.; Li, D.; Liu, H. Be more with less: Hypergraph attention networks for inductive text classification. *arXiv* **2020**, arXiv:2011.00387.
24. Liu, X.; You, X.; Zhang, X.; Wu, J.; Lv, P. Tensor graph convolutional networks for text classification. *Proc. AAAI Conf. Artif. Intell.* **2020**, *34*, 8409–8416. [[CrossRef](#)]
25. Kim, Y. Convolutional neural networks for sentence classification. *arXiv* **2014**, arXiv:1408.5882.
26. Liu, P.; Qiu, X.; Huang, X. Recurrent neural network for text classification with multi-task learning. In Proceedings of the 25th International Joint Conference on Artificial Intelligence, New York, NY, USA, 9–15 July 2016; pp. 2873–2879.
27. Wang, X.; Liu, Y.; Sun, C.J.; Wang, B.; Wang, X. Predicting polarities of tweets by composing word embeddings with long short-term memory. In Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers), Beijing, China, 26–31 July 2015; pp. 1343–1353.
28. Huang, M.; Cao, Y.; Dong, C. Modeling rich contexts for sentiment classification with lstm. *arXiv* **2016**, arXiv:1605.01478.
29. Devlin, J.; Chang, M.W.; Lee, K.; Toutanova, K. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv* **2018**, arXiv:1810.04805.
30. Yang, Z.; Dai, Z.; Yang, Y.; Carbonell, J.; Salakhutdinov, R.R.; Le, Q.V. Generalized autoregressive pretraining for language understanding. *arXiv* **2019**, arXiv:1906.08237.
31. Sun, Y.; Wang, S.; Li, Y.; Feng, S.; Tian, H.; Wu, H.; Wang, H. Ernie 2.0: A continual pre-training framework for language understanding. *Proc. AAAI Conf. Artif. Intell.* **2020**, *34*, 8968–8975. [[CrossRef](#)]
32. Barbosa, L.; Feng, J. Robust sentiment detection on twitter from biased and noisy data. In Proceedings of the 23rd International Conference on Computational Linguistics: Posters, Beijing, China, 23–27 August 2010; pp. 36–44.

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.