J. COMBIN. MATH. COMBIN. COMPUT. 127a (2025) 3263--3278

Journal of Combinatorial Mathematics
and Combinatorial Computing
www.combinatorialpress.com/jcmcc

Research on Optimizing Primary School Classroom Atmosphere Based on Emotional Analysis Model

Yang Li[1,✉]

[1] *School of Teacher Education, Pingdingshan University, Pingdingshan, Henan, 467000, China*

ABSTRACT

In the field of artificial intelligence education, teaching emotion, as the main assessment basis for teaching evaluation, profoundly affects the teaching method, classroom atmosphere and teaching effect of teachers. This thesis proposes a combined network structure, CRNN, by taking advantage of CNN for speech emotion feature extraction and RNN for sequence modeling, and realizes emotion recognition of classroom discourse through DenseNet neural network to realize the crosstalk between each layer and other layers, and LSTM neural network to complete the task of speech emotion classification. On this basis, the open classroom video of the sixth grade of an elementary school is analyzed for sentiment, and the teaching practice of the application of speech emotion recognition model is carried out to study the optimization effect of the model application on the classroom atmosphere of the elementary school. The overall sentiment value of the classroom interaction video floats in the range of 0~1.9, showing a trend of first increasing and then decreasing, reflecting the feasibility of applying the speech emotion recognition model of this paper to classroom sentiment analysis. Through the teaching experiment, the positive emotional performance of the experimental group is more obvious than that of the control group, and 95.46% of the students agree that the application of the model can improve classroom interaction and the overall atmosphere. The speech emotion recognition model studied here can mobilize the classroom atmosphere, and has more important classroom guidance and application significance.

Keywords: CRNN, DenseNet, LSTM, speech emotion recognition, sentiment analysis, classroom atmosphere

1. Introduction

In teaching, a good classroom atmosphere is an important prerequisite for ensuring the smooth progress of teaching [1]. Creating a good classroom atmosphere can improve teaching efficiency, and it is also the key to improve students' learning efficiency [2]. Teachers as the organizer of classroom

✉ Corresponding author.
E-mail address: yangnuo201601@126.com (Y. Li).
Received 15 January 2024; Revised 20 March 2024; Accepted 30 November 2024; Published Online 15 April 2025.
DOI: 10.61091/jcmcc127a-185

teaching, we should pay attention to the combination of the actual situation of students, to create a suitable teaching atmosphere, so as to ensure that students adapt to the teaching atmosphere and are willing to take the initiative to learn, which is also an embodiment of respect for the status of the student body [3]. In creating the classroom atmosphere, only respect the student's subjective position, choose the content suitable for students to learn, in order to mobilize the enthusiasm of students to learn, so that students in the classroom with a full spirit into learning, promote classroom teaching, and then enhance the teaching efficiency as well as to achieve the teaching goals [4-5].

Interest is the best teacher for students to learn, and it is also the psychological tendency to recognize things, which can focus students' attention [6]. Students' interest can be stimulated to obtain good teaching results, so elementary school teachers should pay attention to creating a pleasant, harmonious classroom environment when teaching, and fully display the vitality of the classroom [7]. In this case, teachers should change from the previous lecturer to the leader, from the manager to the guide, for the sake of the students, to create a relaxed and harmonious classroom atmosphere, so that teachers and students become mutual teaching and learning [8].

The implementation of the new curriculum reform has new requirements for elementary school teaching, requiring teachers to pay attention to the cultivation of students' humanistic consciousness in teaching, which requires teachers to create a suitable classroom atmosphere in combination with the actual situation of students. A good teaching atmosphere is the key to promoting students' learning efficiency. In the period when the teaching task is relatively heavy, the use of this approach can effectively cultivate students' love, improve students' moral cultivation and aesthetic interest, and promote the improvement of students' learning efficiency and comprehensive development. Literature [9] explored the factors affecting classroom students' learning emotions in the engineering student teaching classroom, and the study pointed out that the classroom atmosphere driven by the teacher's positivity and enthusiasm as well as the classroom teaching procedures significantly affect students' emotions. Literature [10] empirically analyzed and assessed the performance of the Structured Thinking Activity Tool in the classroom and verified that the tool played a significant value in facilitating teacher-student interactions and fostering students' metacognition. Literature [11] elaborated that community-based interactive atmosphere is the most preferred classroom atmosphere for elementary and middle school students by summarizing elementary school students' views on atmosphere, but there were some differences in perceptions about classroom atmosphere in the two SWAS programs. Literature [12], based on follow-up surveys and in-depth interviews, revealed that school climate is related to children's classroom mood and that positive classroom climate also promotes preschoolers' vocabulary memorization and executive functioning. Literature [13] assessed the level of competence in teaching practices in special education to help optimize and improve the teaching competencies of special education teachers to enhance better quality instruction. Literature [14] elaborated that family climate and family expectations significantly influence children's behavior, including physical activity, and a related teaching experiment conducted in Leipzig, Germany, showed that children's level of physical activity and health were largely in line with family expectations. Literature [15], based on international data surveys, pointed out that the atmosphere of British teaching classrooms was poor and chaotic, and explored the influencing factors behind the atmosphere of the classroom, such as the teacher's personality, the level of teaching in the school, and the characteristics of the students.

The study centers on the classroom teaching scenario, establishes the logical framework of the sentiment analysis model empowering classroom teaching, and provides the theoretical foundation for the subsequent development of the article. Specifically, the study is based on the convolutional neural network DenseNet neural network for speech emotion feature feature extraction, the extracted features are given to the recurrent neural network, and the LSTM neural network in the recurrent neural network is utilized to complete the emotion classification task, so as to obtain the CRNN, a neural network classifier combining the convolutional neural network and the recurrent neural network, and thus realize the speech emotion recognition model. Construction. The classroom video of the sixth grade of elementary school is used as a research sample, and the model is used to analyze the speech emotion of classroom teacher-student interaction, so as to realize the feasibility assessment of the speech emotion recognition model in this paper. We also selected 44 students in an online course of an elementary school as the experimental subjects, divided into experimental group and control group, to carry out the teaching practice of the application of speech emotion recognition model, and explored the effect of the emotion recognition model to enhance the classroom atmosphere by comparing the various emotion recognition results of the experimental group and the control group and investigating the accuracy of the emotion state recognition and the effect of optimizing the classroom atmosphere.

2. Logical framework of the Sentiment Analysis Model to empower classroom teaching and learning

Classroom teaching is a complex process, which is not only about teachers unilaterally imparting knowledge, but also requires interaction and emotional exchange between teachers and students, and personalized classroom teaching is even more demanding. However, the complexity of emotions does not allow teachers and students to perceive and access their emotions easily, which makes it impossible to take appropriate feedback and intervention. In this context, sentiment analysis model, as a technical tool to empower personalized teaching, meets the needs of classroom teaching in terms of supporting emotion recognition, expression and feedback.

Focusing on teaching scenarios, the sentiment analysis model empowers classroom teaching with the collection of multimodal data as the basic premise, multimodal sentiment data analysis as a powerful support, and teaching services as an effective focus point, ultimately serving the needs of personalized teaching scenarios, forming a closed loop of empowerment. Among them, the emotion recognition of teachers and students supported by multimodal data follows the typical computing process from the establishment of data sets to model training and testing, and the basic theory of emotion and artificial intelligence technology standards, information security and privacy protection are the necessary "umbrella" to support the effective application of personalized teaching empowered by affective computing, and together they promote the development of personalized teaching in the direction of temperature and intelligence. Together, they promote the development of personalized teaching in the direction of temperature and intelligence. In this regard, combining the analysis of the needs of personalized teaching with the enabling limits of the sentiment analysis model, and by disassembling and analyzing the core steps in the practice of personalized teaching empowered by the sentiment analysis model, the logical framework of classroom teaching empowered by the sentiment analysis model is shown in Figure 1. The framework is centered on personalized teaching scenarios, guided by the basic theory of emotion and artificial intelligence technology standards, and based on the premise of information security and privacy protection, it realizes the integrated process of emotion analysis model enabling personalized teaching from emotion data collection, storage, modeling and analysis to feedback application, and the modules are closely related to each other, which jointly promotes the realization of a new human-machine symbiosis ecology.

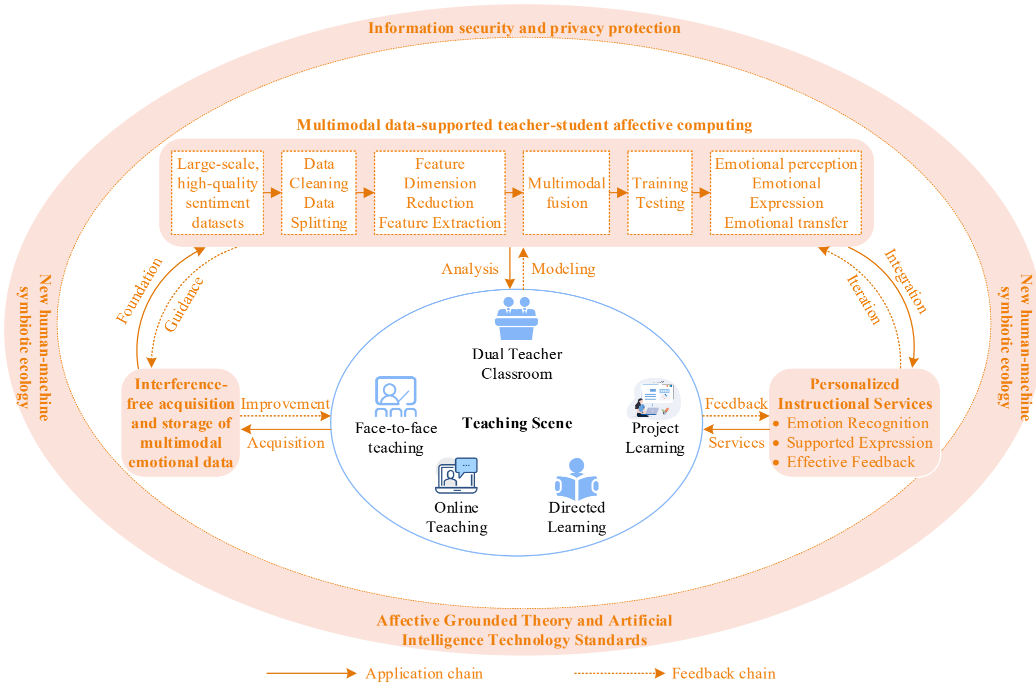


Fig. 1. Emotional analysis of the logical framework of classroom teaching

Overall, this logical framework consists of a coherent application chain and a feedback chain. In the application chain, multimodal affective data are collected and stored from personalized teaching scenarios such as face-to-face teaching, online teaching, self-directed learning, project-based learning, and dual-teacher classrooms based on lightweight and non-disturbing devices, which provide a multi-source data base for teacher-student affective computing. Following that, the data-based, multimodal data-supported teacher-student emotion recognition realizes emotion perception, expression and transfer through deep learning computational flows from data set establishment, data cleaning and splitting, feature extraction, multimodal fusion, model training and testing. Finally, it is applied to the personalized teaching service module to achieve emotion recognition, support expression and effective feedback, and finally serve the personalized teaching scenario. In the feedback chain, based on the quality analysis of the collected multimodal data, the data collection module can feedback the emotion data collection requirements (e.g., modal requirements, quality requirements, scale requirements, etc.) to the teaching scenario, improve the data collection process in the teaching scenario, and realize a higher-quality, high-accuracy, and high-efficiency emotion data collection, and the emotion computation can also reverse guide the emotion data collection from the algorithmic dimension. At the same time, the updated needs of teaching scenarios (e.g., emotional interactions in immersive environments, emotional interactions in informal learning, etc.) will also be fed back to the affective computing module through the personalized teaching service module, so as to continuously update and iterate the teacher-student affective computing model. In addition, all the above processes of empowering classroom teaching with sentiment analysis models should be carried out under the guidance of sentiment-based theories and AI technology standards, as well as under the premise of information security and privacy protection.

3. CRNN-based emotion recognition model for classroom speech

In this paper, we design a network structure CRNN that combines a neural network (CNN) and a recurrent neural network (RNN). This structure takes advantage of both the CNN's advantage for feature extraction and the RNN's advantage of retaining memory capacity on sequence models.This network has good results in both object annotation and speech recognition, and will be used in the classroom for speech emotion recognition.

3.1. CNN and RNN

CNN and RNN are currently one of the most representative algorithms in deep learning.CNN can be used for feature extraction, which uses a convolutional layer to extract localized features by performing convolutional computation on the previous layer of neurons. The purpose is to determine the position between different features by localized features.

In the dimension of three-dimensional or four-dimensional convolution, compared with the ordinary two-dimensional convolution there is one more attribute called channel, using two-dimensional convolution method of calculation, each channel is calculated separately, respectively, multiple channels with multiple convolution kernels for two-dimensional convolution, and finally merged into one channel, the calculation formula is shown in equation (1):

$$Y^{l}(m,n)=X^{k}(m,n)\times H^{kl}(m,n)=\sum_{k=0}^{K-1}\sum_{j=0}^{I-1}\sum_{j=0}^{J-1}X^{k}(m+i,n+i)H^{kl}(i,j) \tag{1}$$

Where, L denotes the output channel of the convolutional layer and K denotes the input channel of the convolutional layer. $K\times L$ denotes the conversion of the number of channels realized by the convolution kernel. X^{k} denotes the features of the kth channel, Y^{L} denotes the features of the L th channel. H^{KL} denotes the convolution kernel in the K th row and L th column.

RNNs are temporal in nature and are mostly used for sequence modeling. Fig. 2 shows the forward propagation of RNN network, first the input $a^{<0>}$ is a zero vector, then the activation value $a^{<1>}$ is computed and then the output is computed $\hat{y}^{<1>}$. The commonly used activation function for RNN networks is the tanh function.

Generally the initial input is the activation value of the zero vector $a^{<0\rangle}$. It is fed into the RNN network.

RNN forward propagation is calculated as in equation (2) and (3):

$$a^{<1>}=g_{1}\left(\omega_{aa}a^{<0>}+\omega_{ax}x^{<1>}+b_{a}\right) \tag{2}$$

$$\hat{y}^{<1>}=g_{2}\left(\omega_{ya}a^{<1>}+b_{y}\right) \tag{3}$$

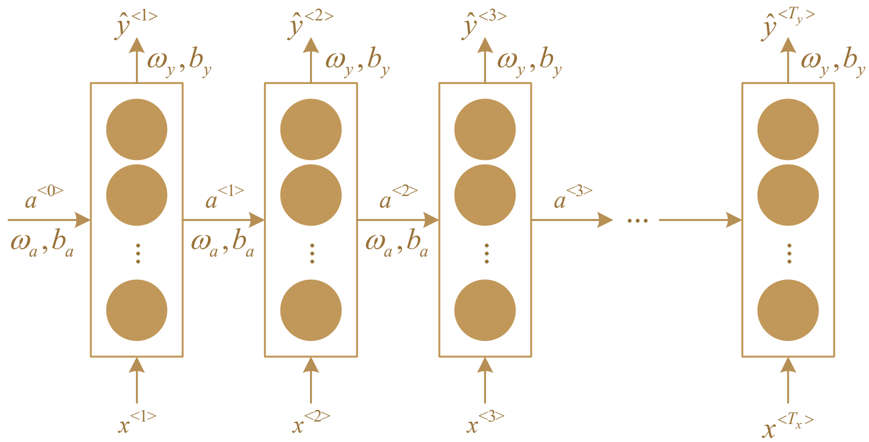


Fig. 2. Cyclic neural network model

More generally, at moment t as in Eqs. (4) and (5):

$$a^{<t>} = g_1\left(\omega_{aa}a^{<t-1>} + \omega_{ax}x^{<t>} + b_a\right) \tag{4}$$

$$\hat{y}^{<t>} = g_2\left(\omega_{ya}a^{<t>} + b_y\right) \tag{5}$$

Where ω_{ax} is the parameter input to the hidden layer. Whereas the activation value is determined by parameter ω_{aa} and the output result is determined by ω_{ya}.

The forward propagation of the RNN network is shown above, proceeding from left to right is forward propagation and the parameters are shared at each time step. In order to build more complex neural networks, the above equation is simplified as in equation (6), equation (7):

$$a^{<t>} = g\left(\omega_a\left[a^{<t-1>}, x^{<t>}\right] + b_a\right) \tag{6}$$

$$\hat{y}^{<t>} = g\left(\omega_y a^{<t>} + b_y\right) \tag{7}$$

Where ω_a is the calculation of a types of quantities that is to calculate the weight of the activation value, b_a for the calculation of the bias of the activation value, ω_y is the calculation of y types of quantities that is to calculate the weight of the output, b_y for the calculation of the bias of the output value.

Compared to the sigmoid function, the Tanh function optimizes the case where the sigmoid function cannot be negative, is more suitable for neural networks, and easily solves the problem of gradient vanishing in neural networks.

In forward propagation, parameters ω_a and b_a are needed in order to compute the activation value $a^{<t>}$, and parameters ω_y and b_y are needed in order to compute the output $\hat{y}^{<t>}$. In back propagation, the loss function $L^{<t>}\left(y^{<t>}, y^{<t>}\right)$ needs to be set up to assign the weights of each time step to the cross-positioned loss function shown in Eq. (8) as an example. In forward propagation, the loss function is the accumulation of the loss function in the whole propagation. The definition is shown in equation (9):

$$L^{<t>}\left(\hat{y}^{<t>}, y^{<t>}\right) = -y^{<t>}\log\hat{y}^{<t>} - \left(1-y^{<t>}\right)\log\left(1-\hat{y}^{<t>}\right) \tag{8}$$

$$L\left(\hat{y}^{<t>}, y^{<t>}\right) = \sum_{t=1}^{T_x} L^{<t>}\left(\hat{y}^{<t>}, y^{<t>}\right) \tag{9}$$

3.2. CRNN Neural Networks

The combination of CNN and RNN is known as CRNN model. Using an end-to-end training strategy avoids greedily forcing the distribution of the intermediate layers to approximate the distribution of the labels and allows for a more appropriate representation from the last layer.

On the one hand, the convolutional layer in CNN focuses on learning the local information of the input. Since each point of the feature map derived from the convolutional layer depends on the product of the convolutional kernel and the corresponding overlapping region of the input. On the other hand, an RNN with a pooling layer can learn the global representation of the input by means of complex nonlinear transformations between time steps. Thus, combining CNNs and RNNs in a speech emotion recognition task helps to model subtle local emotion cues and capture contextual emotion information simultaneously.

The network architecture of an ordinary CRNN speech emotion recognition system consists of three parts, the first two of which form the CRNN.The first part is a convolutional feature extractor which takes as input a spectrogram image representation of a portion of an audio file. This feature extractor convolves the input image and merges it in several steps and produces a spread feature map. Thus, for the case where the speech vocalization is pre-divided into multiple segments, the CNN-learning features are obtained for each segment. Specifically, for a given vocalization denoted as $[s(1), s(2), \ldots, s(T)]$, where $s(t)$ is the segmented spectrogram of the original audio input and T is the number of segments, a sequence of feature vectors $[c(1), c(2), \ldots, c(T)]$ after the CNN module is obtained.

The second part is an RNN, where each time step corresponds to a portion of the original audio input.The RNN can handle inputs of various lengths, so no audio input clipping or padding is applied. An LSTM structure was chosen for the RNN, which is able to address the long term dependencies present in sequential data.The LSTM learns from a sequence of CNN feature vectors $[c(1), c(2), \ldots, c(T)]$ and outputs a sequence $[r(1), r(2), \ldots, r(t)]$. Then, similar to HSF computed on LLDs, the statistics of the LSTM outputs are computed through a pooling layer. In the usual case, only the average pooling is applied. In order to obtain richer statistical information on the output of the LSTM network, they are additionally max-pooled and min-pooled, and the resulting pooled vectors $P_{max}, P_{avg}, P_{min}$ are concatenated into a single vector.The LSTM is set up to have 128 units, and thus the vectors $r(t), P_{max}, P_{avg}, P_{min}$ are all of size 128. The order $r(t)^i$ represents the i th element of $r(t)$. The merging process can be expressed as:

$$P_{\max} = \left(p^{1}_{\max}, \ldots, p^{128}_{\max}\right), p^{i}_{\max} = \max_{1 \le t \le T} r(t)^{t} \tag{10}$$

$$P_{ave} = \sum_{1 \le t \le T} \frac{r(t)}{T} \tag{11}$$

$$P_{\min} = \left(p^{1}_{\min}, \ldots, p^{128}_{\min}\right), p^{i}_{\min} = \min_{1 \le t \le T} r(t)^{i} \tag{12}$$

In addition, the subsequent third part of the network consists of two fully connected layers and a softmax layer that predicts sentiment categories. Their sizes are 128, 32 and 9, respectively.

The following paper describes the two parts that make up the RCNN neural network, DenseNet for the CNN part and LSTM for the RNN part.

3.2.1. DenseNet Neural Networks. The DenseNet neural network is a completely new structure of neural network, which draws on the experience of the ResNet and Inception networks that performed well before it, and proposes a DenseBlock structure that is easier to retain features, allowing the network to be more tightly aligned and allowing the neural network to go deeper, achieving excellent results in most classification tasks.

Convolutional networks can be trained more deeply, efficiently and accurately when they have long jump links between layers that are close to the inputs and outputs. In this paper, we introduce dense convolutional networks, which are used to connect each layer to the other layers through feedforward. And since L layer has L connections: each layer is connected to the other connected layers after it, the network has $L(L+1)/2$ direct connections. By inputting features to all layers before each layer, the feature maps can be input to the subsequent layers.

In the case of a single image in a convolutional network, there is mainly L layer, and in each layer there is a nonlinear transformation $H_\ell(\cdot)$, which indicates the number of layers that can be operated on behalf of the composite function, such as: batch normalization, rectification of linear units, convolutional suppression, or merging, and the output of layer ℓ is represented as x_ℓ.

DenseNet improves the flow of information from layer to layer and proposes different connection patterns, connecting the layers directly to all subsequent layers. Therefore, layer ℓ receives the feature mapping $x_0,\ldots,x_{\ell-1}$ of the previous layers as input:

$$x_\ell = H_\ell\left(\left[x_0, x_1, \ldots, x_{\ell-1}\right]\right) \tag{13}$$

For cascading the feature mappings present in layer $0,\ldots,\ell-1$, because of their tight layer-to-layer connection, this network structure is called a dense convolutional network (DenseNet). To be able to perform the operation, the multiple input cascades present in $H_\ell(\cdot)$ are put into Equation (13) into a tensor.

In this paper, the number of convolutional kernels is set to 32, and a 1x1 convolutional layer is added before each layer of convolutional kernels to ensure that the width of the network is not too large.

Unlike the original DenseNet, the results obtained in this paper after the neural network training can not be directly into the softmax layer for classification, but also need to go through the LSTM neural network to continue training, and finally get the classification results.

3.2.2. LSTM Neural Networks. During RNN training, the information keeps looping back and forth, which leads to a very large update of the neural network model weights, as the accumulation of error gradients during the update will lead to a very unstable network, and in extreme cases, the values of the weights may become so large that they overflow. In addition, the RNN's ability to memorize tensors with longer sequences is weakened, and it does not perform very well in long text sequences.

Long Short-Term Memory Neural Networks (LSTM) are explicitly designed to avoid the long-term dependency problem.LSTM is a modified type of Recurrent Neural Networks (RNN) that learns long-term context-dependent information. And the reason why LSTM is able to solve the long-term dependency problem of RNN is because LSTM introduces the gate mechanism.LSTM contains three gate computations, i.e., forgetting gate, input gate, and output gate. Gates enable selectively letting information through or discarding it, mainly through a sigmoid activation function and a point-by-point multiplication operation. This helps the LSTM to keep track of the information over time.

Suppose the input information is $X_t = \left\{X_1, X_2, \ldots, X_T\right\}$ and after LSTM the output information is $h_t = \left\{h_1, h_2, \ldots, h_T\right\}$. The oblivion gate is denoted by g^f. The oblivion gate decides whether to go over the information or not by judging the importance of the current input information. The inputs are the output of the previous unit h_{t-1} and the input of the current unit x_t, and after the sigmoid layer, the outputs between 0-1 are obtained g^f, which is computed by the formula:

$$g^f = \sigma\left(W^f h_{t-1} + I^f\right) \tag{14}$$

The input gate, denoted by g^i, determines whether the unit state needs to be updated by judging the last output and current input information. Its calculation formula is:

$$g^i = \sigma\left(W^i h_{t-1} + I^i x_t\right) \tag{15}$$

$$g^c = \tanh\left(W^c h_{t-1} + I^c x_t\right) \tag{16}$$

Remember that the formula for updating the state of a cell from m_{t-1} to m_t is:

$$m_t = g^f \odot m_{t-1} + g^i \odot g^c \tag{17}$$

The output gate is denoted by g^o. The output gate is how much the current output depends on the current memory cell state. Its calculation formula is:

$$g^o = \sigma\left(W^o h_{t-1} + I^o x_t\right) \tag{18}$$

$$h_t = g^o \odot \tanh\left(m_t\right) \tag{19}$$

Where h_{t-1} represents the output of the previous cell and x_t represents the input of the current cell. g^i, g^f and g^o are represented as input gate, forgetting gate and output gate respectively. m_t is the current memory cell state value, g^c represents the update value of the candidate memory cell, the sigmoid activation function is denoted by σ, and the activation function for the update value of the candidate memory cell usually uses the tanh function. The element-by-element multiplication is denoted by $\odot$, the weight matrix by W^i, W^f, W^o, W^c, and the projection matrix by I^i, I^f, I^o, I^c.

For this paper, LSTM has excellent feature extraction ability in terms of sequences, the tensor after convolutional modeling is put into the neural network, and the sequences obtained after LSTM will be classified by two layers of fully-connected network and one layer of softmax layer.

4. Analysis of emotions and optimization of the effectiveness of the atmosphere in the elementary school classroom

The chapter first uses the CRNN-based speech emotion recognition model to analyze the verbal interactions in the elementary classroom to verify the effect of the speech emotion recognition model in this paper, and then combines the model with an intervention teaching experiment to analyze its optimization effect on the atmosphere of the elementary classroom.

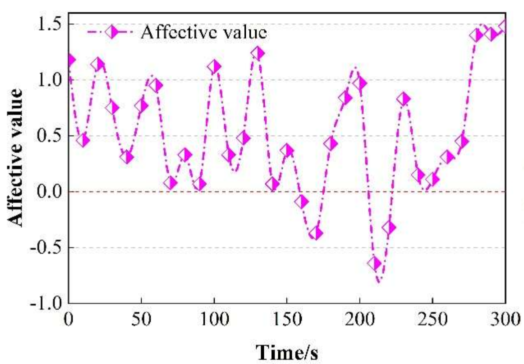
4.1. Speech Sentiment Analysis of Teacher-Student Interaction in the Classroom

4.1.1. Experimental data processing. The classroom video selected for this experiment is a sixth-grade classroom in an elementary school, this classroom video is a public video recorded for offline teaching scenarios, which ensures the authenticity and effectiveness of the experiment, the teacher in the classroom interacts with the students more, and the teacher often allows the students to discuss the problems, the teacher in the video asks many questions for specific problems, the classroom activities mainly include lecturing, questioning, answering, students' practice, discussion, experiments Classroom activities include lecturing, questioning, answering, student exercises, discussions, experiments, etc. By removing ineffective speech segments such as classroom discussions, exercises, experiments, etc., six effective interactive speech segments were retained, and the duration of each classroom speech segment was 300s.

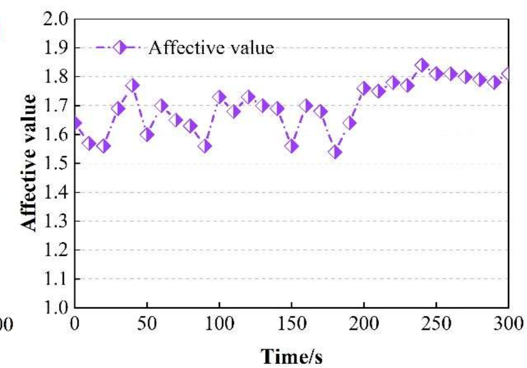
4.1.2. Classroom sentiment analysis. Figure 3 shows the emotion change curves of six classroom interaction segments, where (a) to (f) are the emotion changes of classroom interaction segments 1 to 6, respectively. Classroom interaction segment 1 is the beginning of the classroom, before 150 seconds, the teacher is teaching, the emotion value P is mainly in the range of 0 to 1.5, indicating that the teacher's emotion has not been fully activated at the beginning of the lesson, and is in a low degree of positive emotion. 150 seconds to 250 seconds, the teacher is based on the content of the previous lesson, the teacher is asking the students questions in the classroom, so the overall emotion fluctuates more, and varies in the range of -1 to 1. 250 seconds to 300 seconds, the emotion value P is increasing, indicating that the teacher's emotion is in a low degree of positive emotion. 250 seconds to 300 seconds stage, the sentiment value P is increasing, indicating that the teacher's sentiment is in a positive state.

Classroom interaction segment 2 is mainly about students' explanation of test questions, and the emotion value P is mainly distributed between 1.5 and 1.9, with students' speeches before 200 s. After 200 s, the teacher's emotion changes from the range of 0-1.5 in segment 1 to about 1.8, indicating that students' active speeches have a positive effect on the teacher's emotion, and that there is a correlation between the teacher's emotion and the students' emotion. Comparing with clip one, the emotion value of clip two is significantly higher than clip one, indicating that when the classroom proceeds to 10 minutes, the voice emotion value P reaches a high level and the classroom is already in a fully activated state.

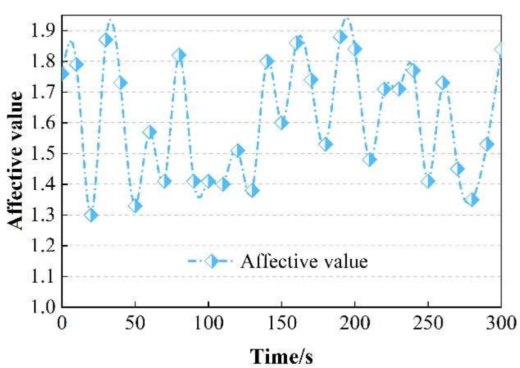
Interactive voice clip three is a practice Q&A clip, the overall affective value P is distributed between 1.3 and 1.9, similar to clip two, the teacher is constantly asking questions, the students' answers are always confident, and the answers are correct, the teacher also gives encouragement and affirmation, indicating that the classroom is always in an active state in the question-answer phase, and the students' emotions are positively mobilized.



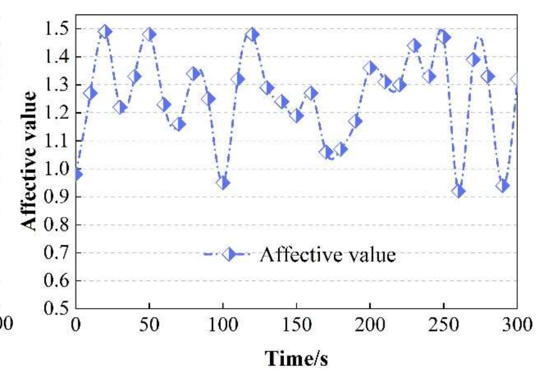
(a)Fragment 1



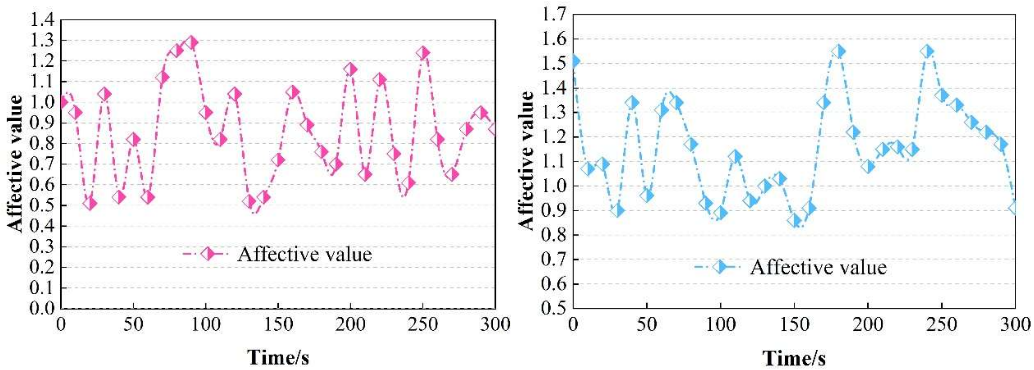
(b)Fragment 2



(c)Fragment 3



(d)Fragment 4



(e)Fragment 5 (f)Fragment 6

Fig. 3. The emotional change curve of six classroom interaction fragments

In classroom interaction clip 4, the students started to explain the reason for the question, which was an open-ended question, with three students taking the initiative to speak, followed by the teacher guiding and communicating with the students according to the question, and the overall classroom affective value P was distributed in the range of 0.9 ~ 1.5. At 100, the teacher gave rewards to the students who performed well, and the voice affective value increased, while at 120 and 150 seconds, the teacher asked the students in a questioning tone, and the voice pleasantness decreased to near 1.2, indicating some validity of the model.

Classroom interaction clip five is mainly for the accompanying practice of answering questions, with students explaining the topics at the blackboard, because it is explained by three different students, each of whom has a different level of emotional pleasantness, and there may be nervousness in the blackboard explanations, so the emotions show a large fluctuation. For clip five, because the P distribution of emotion value is 0.5 ~ 1.3, it presents a significant decrease in the emotion of pleasure compared to the previous one, which indicates that the students' motivation in the classroom decreases in the later part of the classroom.

Classroom interaction clip six is the classroom summary part, the teacher asked students to think about the learning gains of this class, two students share learning gains, classroom emotion also shows a higher trend, P value distribution in 0.8 ~ 1.6, indicating that the free speech state, students' emotions are higher.

Analyzed as a whole, the distribution of the emotion values of the classroom interaction segments were 0 ~ 1.5, 1.5 ~ 1.9, 1.3 ~ 1.9, 0.9 ~ 1.5, 0.5 ~ 1.3, 0.8 ~ 1.6, and the emotion of classroom interaction pleasure showed a trend of increasing and then decreasing. From the above analysis, it shows that this classroom teaching atmosphere is good, and the teacher's teaching methods are effective, enriching the classroom through discussions, exercises, experiments and other forms, stimulating students' enthusiasm for the class, so the classroom pleasantness is sustained for a long time, and the teacher makes good use of rewards and encouragement to drive students to think positively.

4.2. Analysis of the effect of classroom atmosphere optimization

In this section, a classroom speech emotion recognition model based on CRNN is practiced for teaching intervention applications using realistic online learning as an application environment.

4.2.1. Selection of experimental subjects. This paper randomly selected a common primary school as the study range, randomly selected 22 students from both the fifth and sixth grade, and ensured that the students' performance distribution was consistent with the overall sample. The above 44 students were divided into the experimental group and the control group to ensure that the grades and performance of the two groups were consistent. During the learning process, the experimental group adopts the classroom speech emotion recognition model designed in this study to assist teaching, and the control group learners adopt the traditional CNN-based speech emotion recognition method. In particular, the traditional voice emotion recognition method first adopts the full-connected network filtering feature, and then as the input of CNN, thus output recognition results. Both groups of learners were given a knowledge quiz at the end of the study, and the corresponding experimental effects were analyzed according to the end of the experiment.

4.2.2. Experimental program:

1) Segmentation of Emotional Polarity. The teacher and student emotion recognition data constructed in this study contains seven categories of emotion state discrete categories of concentration, pleasure, surprise, confusion, frustration and boredom. In order to improve the flexibility and convenience of the emotion state categories in real-world scenarios, this study divides the above seven categories into three emotion polarities, i.e., positive (focused, pleasant), neutral (surprised), and negative emotions (confused, frustrated, and bored).

2) Types of learners. Learners with different types of affective responses were graded, with type I being excellent learners, meaning students with a high proportion of positive affective states in the learning process and a concentrated distribution of focused emotions, which can promote a high degree of participation in the learning process with a high degree of focused emotional distribution, and do not require frequent intervention services. type II is a potentially risky learner, with a low proportion of negative affective states (10% to 20%), and requires frequent intervention services. 10% to 20%), require a slightly more intense intervention than Type I, and can be guided to reflect on their learning and consolidate their learning activities accordingly. Type III high-risk learners, which refers to learners with a high percentage of negative affect (more than 20%) or a concentrated distribution of negative affect, whose learning process is dominated by negative affect, should be given a high level of individual targeted intervention.

4.2.3. Analysis of application effects. The length of the learning content of this experiment is 30 minutes, using every 5 seconds of learning emotional state recognition and recording, so the total number of emotional state recognition of each subject learner is supposed to be 360.Table 1 shows the results of emotional state recognition of the learners in the experimental group. Table 2 shows the emotional state recognition results of the learners in the control group. It is obvious that the number of emotional state perception missing for the three types of learners in the experimental group is much lower than that of the control group, and the CRNN classroom speech emotion recognition model proposed in this study is much more robust than the traditional CNN-based speech recognition method in real online learning scenarios.

The number of recognized pleasant emotions of the three types of students in the experimental group is 42.86% ~ 54.55% higher than that of the control group, the number of recognized focused emotions is 11.07% ~ 17.67% higher than that of the control group, and in terms of test scores, the experimental group is 4.16% ~ 9.02% higher than that of the control group, which suggests that the positive classroom emotions of the students in the experimental group are more pronounced, and the atmosphere in the classroom is better.

Table 1. The emotional status recognition results of learners in the experimental group

	Type I	Type II	Type III
Pleasure	17	10	11
Focused	321	279	233
Surprised	5	9	12
Confused	7	35	57

Angry	0	0	2
Despondent	1	6	13
Weary	1	15	25
Missing	8	6	7
Total	352	354	353
Test score	92.63	82.74	71.66

Table 2. The emotional status recognition results of learners in the control group

	Type I	Type II	Type III
Pleasure	11	7	8
Focused	289	243	198
Surprised	7	5	9
Confused	12	41	62
Angry	2	2	6
Despondent	3	6	8
Weary	5	8	14
Missing	31	48	55
Total	329	312	305
Test score	88.93	77.65	65.73

Changes in learners' successive affective states were analyzed, and Figure 4 shows the changes in affective states in the control and experimental groups after the intervention, where the vertical coordinate values of -4 to 3 corresponded to anger, boredom, frustration, confusion, perceptual lack, surprise, pleasure and concentration in that order. The affective state of the control group students is greater than 0 (concentration and pleasure) in the first 20s, and only a short period of positivity occurs, while the affective state of the experimental group students is greater than 0 (concentration and pleasure) in the first 50s, and on the whole, most of the succeeding emotions of the experimental group students are positive emotions, and the application of the CRNN classroom speech emotion recognition model makes the experimental group's intervention effect more considerable.

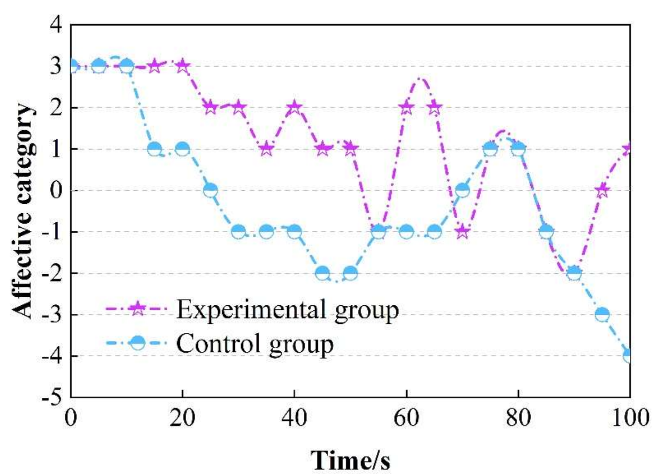


Fig. 4. Emotional state changes after the intervention

4.2.4. Student Experience Survey. To further verify the effectiveness of the proposed CRNN classroom speech emotion recognition method in real learning environments, this study designed a questionnaire "Survey on the Experience of Using the CRNN Classroom Speech Emotion Recognition Model" in the application practice, and the following results are the results of the questionnaire data collation and analysis.

1) Accuracy of Emotion State Recognition. The purpose of the survey on the dimension of emotion state recognition accuracy is to determine, from the perspective of the learners, whether the emotion state perception results of the emotion recognition method proposed in this study coincide with the learners' emotion states at the corresponding moments in the learning process, so as to indirectly prove the performance of the method in terms of emotion recognition accuracy. Table 3 shows the different categories of learners' perceptions of emotion recognition accuracy. On the whole, 77.28% of the learners in the experimental group think that the feedback of the platform's emotion recognition information is in line with their own emotional state of learning at that moment, while only 2 learners do not quite agree with the feedback results of the platform's emotion recognition function, and 3 learners are not sure. Type I as well as Type II learners recognized the emotion recognition function more, with 83.33% and 77.78% respectively. Overall, the results of the questionnaire in this dimension validate the accuracy of the emotion recognition method proposed in this study through the learner perspective attitude.

Table 3. Different categories of learners' views on the accuracy of emotional recognition

	Type I	Type II	Type III	Total
Very accurate	2	1	2	5
	33.33%	11.11%	28.57%	22.73%
Comparably accuracy	3	6	3	12
	50.00%	66.67%	42.86%	54.55%
Indeterminate	1	1	1	3
	16.67%	11.11%	14.29%	13.64%
Very inaccurate	0	1	1	2
	0.00%	11.11%	14.29%	9.09%
Comparably inaccuracy	0	0	0	0
	0.00%	0.00%	0.00%	0.00%

2) Classroom atmosphere optimization effect. The results of the questionnaire survey on the optimization effect of classroom atmosphere are shown in Table 4. The percentage of students who think that the model application can improve learning enthusiasm, classroom interaction and overall atmosphere are 86.36%, 95.46% and 95.46% respectively, and the CRNN classroom speech emotion recognition model can effectively stimulate students' enthusiasm and optimize the classroom atmosphere.

Table 4. Results of questionnaire survey of classroom atmosphere optimization effect

	Improve learning initiative	Improve classroom interaction	Improve the overall atmosphere
Very accurate	6	7	10
	27.27%	31.82%	45.45%
Comparably accuracy	13	14	11
	59.09%	63.64%	50.00%
Indeterminate	2	1	1
	9.09%	4.55%	4.55%
Very inaccurate	1	0	0
	4.55%	0.00%	0.00%
Comparably inaccuracy	0	0	0
	0.00%	0.00%	0.00%

5. Conclusion

This study combines CNN and RNN neural networks to propose a speech emotion recognition model based on CRNN. It also selects six clips from the public classroom video of a sixth grade elementary school for classroom sentiment analysis, and selects 44 students from an online course of an elementary school as the experimental subjects to verify the classroom atmosphere optimization results of the speech emotion model in this paper through teaching experiments. The research results are as follows:

1) The emotion values of the six classroom interaction segments are 0 ~ 1.5, 1.5 ~ 1.9, 1.3 ~ 1.9, 0.9 ~ 1.5, 0.5 ~ 1.3, 0.8 ~ 1.6, respectively, and the overall emotion values show a trend of rising and then dropping, which verifies the effect of speech emotion recognition of this paper's model, and also indicates that the sample classroom has a better teaching atmosphere, and the interaction between the teachers and students shows a longer period of positive emotions.

2) Through the teaching practice of speech emotion recognition model application, the pleasure emotion recognition and concentration emotion recognition of the students in the experimental group are higher than that of the control group by more than 42.86% and 11.07%, respectively, and the test scores are higher by 4.16% ~ 9.02%.86.36% of the students believe that the model can improve the learning motivation, and 95.46% believe that it can enhance the classroom interaction and the overall atmosphere of the classroom. The optimization effect of the speech emotion recognition model on the classroom atmosphere is confirmed.

3) Speech emotion recognition has a very important application in classroom scenarios, which helps to trace back the classroom behavior itself from the perspective of emotion assessment, so as to guide the classroom activities correctly. The emotion recognition method proposed in this study focuses on stable and efficient emotion recognition from the speech dimension, and in the future, other features can be further considered to realize a multimodal robust model as well as to consider the construction of temporal features.

References

[1] M. H. M. BeldD. van der VoortG. H. P. van der HelmC. H. Z. KuiperJ. J. W. de SwartG. J. J. M. Stams. (2018). Assessing classroom climate in special education: a validation study of the special education classroom climate inventory. Journal of psychoeducational assessment, 36(7).

[2] Mira-Galva, M. J. , & Gilar-Cobi, R. . (2021). Okapi, an emotional education and classroom climate improvement program based on cooperative learning: design, implementation, and evaluation. Sustainability, 13.

[3] Peter, B. , & Rob, W. . (2018). Classroom contexts for learning at primary and secondary school: class size, groupings, interactions and special educational needs. British Educational Research Journal, 44(4), 681-703.

[4] Sina Schürer, Ophuysen, S. V. , & Behrmann, L. . (2020). A new instrument for assessing cohesion in primary and grammar school classes—factorial structure and measurement invariance of the gruko4. Journal of Psychoeducational Assessment.

[5] Sipone, S. , Víctor Abella-García, Rojo, M. , & Dell'Olio, L. . (2021). Using classcraft to improve primary school students' knowledge and interest in sustainable mobility. Sustainability, 13.

[6] Liu, R. , Wang, L. , Koszalka, T. A. , & Wan, K. . (2022). Effects of immersive virtual reality classrooms on students' academic achievement, motivation and cognitive load in science lessons. Journal of computer assisted learning.

[7] Kikas, M. K. . (2018). Associations between reading skills, interest in reading, and teaching practices in

first grade. Journal of neurosurgical sciences, 62(6).

[8] Aidinopoulou V D G . (2017). An action research study from implementing the flipped classroom model in primary school history teaching and learning. Journal of Educational Technology & Society, 20.

[9] Hartikainen, S. , Pylvs, L. , & Nokelainen, P. . (2022). Engineering students' perceptions of teaching: teacher-created atmosphere and teaching procedures as triggers of student emotions. European journal of engineering education.

[10] Branigan, H. E. , & Donaldson, D. I. . (2019). Learning from learning logs: a case study of metacognition in the primary school classroom. British Educational Research Journal, 45(2).

[11] Farmer, J. L. , Spearman, M. , Qian, M. , Leonard, A. E. , & Rosenblith, S. . (2018). Using children's drawings to examine student perspectives of classroom climate in a school-within-a-school elementary school. The Elementary School Journal, 118(3), 000-000.

[12] Rochester, S. E. , Weiland, C. , Unterman, R. , Mccormick, M. , & Moffett, L. . (2019). The little kids down the hall: associations between school climate, pre-k classroom quality, and pre-k children's gains in receptive vocabulary and executive function. Early Childhood Research Quarterly, 48, 84-97.

[13] Rodríguez-Dorta Manuela, & Borges áfrica. (2017). Behavioral patterns in special education. good teaching practices. Frontiers in Psychology, 8(31).

[14] Alexandra, Z. , Christina, N. , Andreas, S. , Hagen, W. , Heike, S. , & Petra, W. . (2022). O3-5family health climate, parental outcome expectations and physical activity of children in primary-school age. European Journal of Public Health(Supplement_2), Supplement_2.

[15] Jenkins, A. , & Ueno, A. . (2017). Classroom disciplinary climate in secondary schools in england: what is the real picture?. British Educational Research Journal.