International Journal for Housing Science and Its Applications
Publish August 10, 2025. Volume 46, Issue 4 Pages 2585-2597

https://doi.org/10.70517/ijhsa464217

Theoretical Basis and Practical Exploration of Intelligent Analysis Methods for Teaching Behavior Data in Higher Education Institutions in the Era of Big Data

Dan Zhang[1,*]
[1] Social Training College, Jilin Open University, Changchun, Jilin, 130000, China
Corresponding authors: (e-mail: zhangdanjilin@163.com).

Abstract In the era of big data, the analysis of classroom teaching behavior in higher education institutions is increasingly becoming automated, informatized, and intelligent. This study investigates methods for analyzing teaching behavior data in higher education institutions and proposes a classroom speech emotion recognition model based on multi-feature fusion. Residual networks and LSTM networks are used for deep feature extraction, while the encoder part of the Transformer is employed for feature fusion. Through experiments on the dataset, the language emotion recognition accuracy of the model in different datasets was below 85%, demonstrating the accuracy of the proposed method for speech emotion recognition. Additionally, the recognition accuracy for each emotion was 6.63% to 17.17% and 16.50% to 20.44% higher than that of the comparison methods. Analysis of speech sentiment in real-world teaching interactions revealed that pleasant emotions in classroom interactions exhibit a trend of first increasing and then decreasing. The sentiment values of interaction segments are sequentially [-1, 1.25], [1.5, 2.0], [1.2, 2.0], [0.85, 1.3], [0.4, 1.25], and [0.8, 1.4], respectively, validating the rationality of the proposed method. It can serve as an intelligent analysis method for teaching behavior data in higher education, assisting teachers in obtaining classroom feedback and optimizing teaching quality.

Index Terms teaching behavior analysis, residual network, LSTM, Transformer, language sentiment recognition

I. Introduction

With the increasing application of artificial intelligence in education and changes in classroom teaching environments, the analysis of student teaching behaviors has become a key focus of educational research, as well as an important means of promoting teacher professional development and educational reform [1]-[3]. Classroom teaching is a crucial activity in China's current basic education system, serving to promote students' comprehensive development from the outset of teaching activities [4]. Teaching activities are composed of a series of teaching behaviors, forming a behavioral system constituted by the interactions between teachers and students [5]. At present, the analysis of classroom teaching behaviors in higher education institutions is gradually transitioning from traditional classroom observation methods to the era of big data. By employing big data analysis techniques to conduct in-depth analyses of classroom teaching behaviors, it is possible to accurately identify teaching issues and establish a new data-driven educational governance model, thereby injecting new momentum into educational innovation [6]-[8].

The classroom serves as a vital base for teaching and learning research and is the primary venue for educational activities [9]. In classroom teaching, traditional observation and analysis methods include manual recording of observations and audio-visual monitoring, which are labor-intensive and prone to errors [10]-[12]. With the emergence of big data analysis technology, intelligent systems can automatically analyze and mine data from classroom videos and audio recordings using analytical indicators, replacing the cumbersome manual analysis with intelligent machine analysis [13]-[15]. With the application of data mining technology, intelligent technical means can be used to rapidly analyze classroom behavior. We will no longer be limited to the manual analysis of professional teachers or experts, but can conduct automatic analysis on a large scale, providing more personalized, targeted, and precise high-quality services for classroom teaching behavior analysis [16]-[19]. Intelligent and rapid analysis will promote the deep integration of data mining and teaching, enabling in-depth research into the application of big data analysis in optimizing teachers' classroom behaviors, and having a profound impact on empowering teachers' lifelong learning and development.

Traditional classroom teaching behavior analysis is primarily based on the Flanders Interaction System (FIAS) for analyzing classroom teaching behavior. Sainyakit, P., and Santoso, Y. I. employed the FIACS analysis method

to examine teaching interaction processes in the classroom, focusing on teacher-student and student-student interaction processes. By combining observational data with statistical calculations, they clarified the interactive logic within classroom teaching [20]. Li, H., and Zeng, X. combined the iFIAS coding system to analyze teacher-student interaction patterns in secondary school English classrooms. The generated video observation matrix table clearly reflects indicators such as student engagement, providing a reference for improving classroom interaction effectiveness [21]. Fang, Z., et al. integrated the Online-Offline Comparative Analysis System (OOCP) into traditional FIAS analysis methods, incorporating internet-based presentations, chat room discussions, and other interactive forms to enhance the effectiveness of analysis of blended online-offline classroom interactions [22]. Wang, D. proposed an improved OOTIAS coding system based on the Flanders Interaction Analysis System (FIAS) and the Information Technology Interaction Analysis System (ITIAS), which possesses the ability to conduct in-depth analysis of smart classroom interaction behaviors and demonstrates excellent human-computer interaction effects [23]. However, with the advent of the big data era and the widespread adoption of smart classrooms, traditional teaching interaction behavior analysis methods have become increasingly inadequate for analyzing teaching behaviors in information-based teaching environments.

In the intelligent era, many scholars have begun to explore the application of automated analysis technologies in classroom settings, conducting comprehensive automated analyses across multiple data dimensions such as behavior and emotion. Wang, W. investigated the performance of the YOLOv5 deep learning algorithm in analyzing teaching behaviors in art classrooms, automatically identifying and classifying students' classroom behavior patterns, thereby providing an effective approach to further enhance student learning quality [24]. Shi, S., et al. utilized artificial intelligence analysis tools to analyze teaching behavior data related to teacher expression and classroom atmosphere. The obtained teacher expression recognition results and dynamic changes in classroom atmosphere data provided data support for optimizing classroom teaching design [25]. Jasim, A. H., and Hoomod, H. K. proposed using hybrid deep learning technology to analyze video data containing student classroom behavior, thereby optimizing interactive patterns in teaching to create rich emotional support and learning experiences for students [26]. Jia, Q. and He, J. constructed a smart classroom behavior analysis model integrating YOLOv5, attention mechanisms, and OpenPose behavior detection, enabling accurate student behavior analysis in complex teaching environments and providing effective solutions for optimizing academic management [27]. Chen, G. and Zhou, J. established a student behavior prediction and analysis model based on convolutional neural networks (CNNs), which accurately predicts students' learning states and behavioral trends in classroom environments, providing decision support for educational administrators in designing teaching plans [28]. Zou, X. designed a parallel computing fruit fly optimization-based adjustable recurrent neural network (PFFO-ARNN) and applied it to online classroom teaching management, achieving more generalized and precise student behavior data analysis [29]. Gong, B. and Jing, F. explored a student intelligent classroom behavior recognition method based on the random forest algorithm and corrected matrix. The proposed model can promptly provide educational administrators with data feedback on students' perceptions of teaching and teaching activities, thereby supporting educational reform [30]. Integrating deep learning and artificial intelligence technologies into classroom teaching behavior analysis can assist teachers in intelligently analyzing and evaluating classroom teaching behavior.

In fact, student behavior data reflecting classroom teaching quality is not limited to the classroom. Existing educational management platforms contain vast amounts of information, including data on student learning habits and course preferences. Leveraging big data technology to further uncover the potential value of student behavior data is a key driver for exploring and implementing reforms and transformations in higher education.

This paper explores data-driven intelligent analysis of teaching behavior in higher education institutions. It selects MFCC and LPC features for speech feature extraction in classroom settings and employs statistical features and similarity matrices to supplement and optimize these features. Subsequently, the obtained MFCC spectrograms and time-series LPC features are subjected to deep feature extraction using ResNt50 and bidirectional long short-term memory networks. A Transformer encoder is then utilized to fuse multiple features, thereby constructing a classroom speech emotion recognition model. Model experiments are conducted on the CASIA and Emo-DB datasets to compare the recognition accuracy of the proposed method with other speech emotion recognition methods, as well as the recognition performance of different methods for various emotions. Based on this, a speech emotion dataset for actual classrooms is constructed. After completing data processing and emotion data annotation, multiple interaction segments from a specific classroom are used as examples to analyze the emotional trend changes in that classroom and the emotional changes in classrooms with different ratings, thereby achieving the practical exploration of the speech emotion recognition method proposed in this paper in real classroom scenarios.

II. Intelligent analysis of teaching behavior data in higher education institutions

The actual learning process is often complex and multifaceted. The instructional behavior data generated in the classroom encompasses not only teacher behavior, student behavior, and teacher-student interaction behavior, but also involves classroom instructional content, classroom instructional context, and changes in teacher-student emotional dynamics. Artificial intelligence technology is increasingly integrating into everyday real-world teaching, making the acquisition of classroom instructional behavior analysis data more convenient, automated, and multi-sourced, thereby highlighting the value and significance of classroom instructional behavior analysis research. For large-scale, multimodal classroom behavior data, researchers can employ appropriate machine learning algorithms to mine the data, convert behavioral information into data, and explore underlying behavioral patterns, trends, and habits. This facilitates researchers' analysis of the learning process, understanding of learning outcomes, and optimization of the learning environment. The deep integration of artificial intelligence technology and education has driven the progress of educational intelligence and achieved significant results. For example, intelligent teaching systems utilizing natural language can process and analyze learners' non-verbal communication patterns, enhancing human-machine interaction. Adaptive learning systems employing artificial intelligence, multi-modal big data, and other technologies can track learners' learning status and content in real time, intelligently adjust learning methods, and implement precise teaching to promote their personalized development. Thanks to 5G-enabled smart virtual reality, real-time VR/AR/MR and remote learning are better supported, enabling personalized and contextualized education. New smart classrooms built using technologies such as the Internet of Things, cloud computing, wearable devices, and artificial intelligence enable more automated and intelligent collection and analysis of classroom teaching behavior data. An artificial intelligence-supported classroom teaching analysis framework can also be constructed, aiming to utilize artificial intelligence technology to standardize, streamline, quantify, and scale classroom teaching behavior analysis.

As AI technology penetrates the education sector, various intelligent technologies are being integrated into classroom teaching, facilitating the acquisition, processing, and analysis of classroom teaching behavior data, enhancing researchers' efficiency, and making it possible to achieve automated, scalable, and routine classroom teaching behavior analysis. This provides strong support for optimizing teaching quality and strategies.

III. Intelligent analysis methods for teaching behavior data

By analyzing teaching behavior data in higher education institutions using artificial intelligence technology, precise data references can be provided for teachers' teaching reflection and professional development. Based on this, this chapter primarily explores intelligent analysis methods for teaching behavior data in higher education institutions, proposes a classroom speech emotion recognition model based on multi-feature fusion, and conducts experimental evaluations of it.

III. A. Feature Selection

III. A. 1) MFCC Features

The Mel spectrum is a representation of the short-term energy spectrum of sound, based on a linear transformation of the logarithmic power spectrum of nonlinear Mel frequencies. Mel frequency cepstral coefficients (MFCC) utilize the Mel scale, which aligns with the response of the human auditory system. Currently, MFCC is widely used in speech recognition systems. In MFCC feature extraction, the same processing steps are shared with LPC, such as preprocessing, framing, and windowing. After windowing, a fast Fourier transform (FFT) is performed, followed by Mel-space filtering to obtain the Mel spectrum. The FFT primarily transforms the audio signal from the time domain to the frequency domain. The application of the Mel scale is primarily because human perception of sound frequency is not linear. This discovery necessitates that acoustic feature extraction undergo a set of non-uniformly spaced Mel scale filters. The Mel scale exhibits exponential growth beyond 1 kHz but is nearly linear below 1 kHz. Formula (1) defines the Mel scale:

$$Mel(f) = 2595 * \log 10(1 + f / 700) \tag{1}$$

Convert the signal to the spectrogram domain using natural logarithm operations, perform feature decorrelation using DCT transformation, and rank them in descending order according to the amount of speech signal information they contain.

III. A. 2) LPC Features

Based on the process of speech formation, phonemes can be regarded as the result of the source excitation signal being influenced by different shapes of the vocal tract. Generally, this is based on the assumption that the source model and the vocal tract model are independent of each other. Linear prediction techniques derive filter coefficients (corresponding to the vocal tract) by minimizing the mean squared error between the input and

estimated samples. The specific calculation methods for these coefficients involve autocorrelation or covariance methods, as shown in Formula (2), which represents the full pole form of the vocal tract transfer function:

$$H(z)=\frac{G}{A(z)}=\frac{G}{\left(1+a_1z^{-1}+a_2z^{-2}+\cdots+a_pz^{-p}\right)} \tag{2}$$

The value of a_i is the LPC coefficient, and G represents the gain or amplitude associated with the vocal tract excitation.

In the extraction of LPC and MFCC features, preprocessing, framing, and windowing are all performed in the same manner. Normalization, average difference, and subsequent pre-emphasis are the primary steps in purifying the speech signal. Pre-emphasis processing is applied to the digitized signal, with the pre-emphasis filter used to smooth the spectrum and mitigate the effects of limited precision. Due to mismatched training and testing conditions, portions of the data that do not carry important information can be directly filtered out. For example, volume differences between different recording devices are irrelevant to recognition. To reduce the impact of such irrelevant factors, normalization is required. During normalization, each sample value of the speech signal is divided by the highest amplitude value in the sample. The DC offset is removed by subtracting the average value of the speech signal from the signal. Pre-emphasis processing involves dividing the speech samples into overlapping frames to minimize discontinuities at the beginning and end of each frame. The Hamming window is smoother than other window functions, as shown in Formula (3), which is the Hamming window function:

$$W(n)=0.54-0.46*\cos\left(\frac{2n\pi}{N-1}\right) \tag{3}$$

Among them, $0\le n\le N-1$. After windowing, the signal undergoes autocorrelation processing, and the highest autocorrelation value p is the order of the LPC analysis. Typically, p is set between 8 and 16. The LPC analysis outputs the $p+1$ autocorrelation coefficients of each frame to the LPC coefficient set.

III. A. 3) Supplementation and optimization of features

(1) Statistical features

Although MFCC and LPC features each have their own advantages and can effectively characterize emotions, However, since the distribution of emotional information in speech signals is not uniform or continuous, statistical features that reflect changes in the signal waveform can serve as a useful supplement. A set of statistical features effective for speech emotion recognition tasks was selected: duration, maximum value, minimum value, mean, standard deviation, root mean square, peak-to-peak value, skewness, kurtosis, waveform factor, peak factor, margin factor, pulse factor, zero-crossing rate, and short-term energy.

(2) Similarity matrix

The extracted LPC feature vectors are processed using the principal component analysis algorithm to extract and retain the parts that have a greater impact on the emotion recognition results.

Among them, principal component analysis (PCA) is an unsupervised dimension reduction algorithm. The algorithm steps are as follows, assuming that there are m lines of n-dimensional original data:

1) Organize the original data into an n-row, m-column matrix, denoted as X.

2) Perform zero mean calculation on each row of X.

3) Obtain the covariance matrix $C=\frac{1}{m}XX^T$.

4) Calculate the eigenvalues and eigenvectors of the covariance matrix.

5) Arrange the eigenvectors by row according to the size of the eigenvalues, and take the first k rows as the matrix P.

6) The final reduced-dimension data is $Y=PX$.

The reduction in dimension results in only k dimensions being retained, and the remaining eigenvectors corresponding to the eigenvalues are discarded. Since the new feature vectors cannot be determined before the new feature matrix is generated, and the new feature matrix is not readable after generation, it is impossible to determine which features from the original data the new feature matrix is composed of. Although the new features still contain information from the original data, their meaning is different from the original. Therefore, dimension reduction algorithms such as PCA are also a form of feature extraction.

III. B. Speech Emotion Recognition Model

The speech emotion recognition model architecture proposed in this paper is shown in Figure 1. Residual networks and bidirectional long short-term memory networks are used to perform deep extraction on MFCC spectrograms

and time-series LPC features, respectively, and the Transformer encoder is used for feature fusion. The classifier at the rear is used to predict the probability of emotion categories.

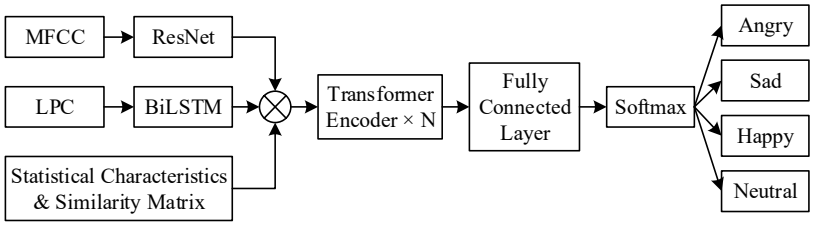


Figure 1: The structure of the speech emotion recognition model

III. B. 1) Residual Network

Residual networks are used to solve image recognition problems. The deeper the network, the more convolutional layers and parameters it has, and the more computational resources and time are required during training and inference.

Assuming that the l th residual unit inputs image features x_l and outputs features x_{l+1}, the residual unit is:

$$y_l = h(x_l) + F(x_l, \omega_l) \tag{4}$$

$$x_{l+1} = \mathrm{Re}\,LU(y_l) \tag{5}$$

When y_l and x_l have the same dimension, $h(x_l)$ is the identity mapping, i.e., $h(x_l) = x_l$. When y_l and x_l have different dimensions, $h(x_l)$ is a linear mapping of x_l, i.e., $h(x_l) = \lambda \cdot x_l$, to match the dimension, $F(x_l, \omega_l)$ is the residual function:

$$F(x_l, \omega_l) = \mathrm{ReLU}(x_l \cdot \omega_l + b_l) \tag{6}$$

In the equation: ω_l is the weight and bias of the l th residual unit, $\omega_l = \{w_{l,k} \mid 1 \le k \le K\}$ where K is the number of layers in the residual unit network, and b_l is the bias of that layer. The activation function is the rectified linear unit (ReLU):

$$\mathrm{Re}\,LU = MAX(0, x) \tag{7}$$

Assuming the loss function is ε, and the input feature of the L th residual unit in any deeper layer is x_L, then the gradient of the residual unit is:

$$\frac{\partial \varepsilon}{\partial x_l} = \frac{\partial \varepsilon}{\partial x_L} \cdot \frac{\partial x_L}{\partial x_l} = \frac{\partial \varepsilon}{\partial x_L}\left(1 + \frac{\partial}{\partial x_l}\sum_{i=1}^{L-1} F(x_i, \omega_i)\right) \tag{8}$$

In the equation, the gradient $\frac{\partial \varepsilon}{\partial x_l}$ can be decomposed into two parts. $\frac{\partial \varepsilon}{\partial x_L}$ indicates that information does not propagate directly through the weight layer, allowing the loss function information to be propagated backward to any shallower unit. At the same time, it ensures that even if the weights are arbitrarily small, the vanishing gradient phenomenon can be avoided.

Among these, ResNet50 has relatively low computational cost, so this structure was selected in the experiment for deep extraction of MFCC spectrograms. The 50-layer convolutional neural network of ResNet50 enables it to learn more complex features, thereby improving accuracy. Additionally, ResNet50 employs residual blocks. Each residual block consists of two convolutional layers and one skip connection. The skip connection directly transmits the input to the output, effectively addressing the vanishing gradient problem. The residual block structure is shown in Figure 2.

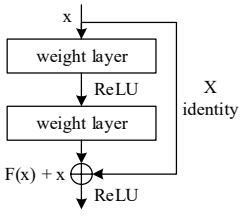


Figure 2: Residual block structure

In the figure, x represents the input, the weight layer region is the convolution layer, and the residual mapping part is represented by $F(x)$. The activation function used is ReLU. Additionally, ResN50e employs a global average pooling layer, which effectively reduces the number of model parameters, thereby mitigating the risk of overfitting.

III. B. 2) Bidirectional Long Short-Term Memory Network

The LSTM network recurrent unit structure consists of three gates: the forget gate f_t, the input gate i_t, and the output gate o_t. The forget gate operates as follows:

$$f_t = \sigma\left(w_{xf} \cdot x + w_{hf} \cdot h_{t-1} + b_f\right) \tag{9}$$

In the equation: w_{xf} and w_{hf} are the weight coefficients of the forget gate, and b_f is the bias term of the forget gate. The output f_t of the forget gate is an n-dimensional output, with each value between $(0,1)$. Information with values close to 0 is forgotten, while information with values close to 1 is retained.

Therefore, through the forget gate, LSTM can remember important information for a long time, and the memory can be dynamically adjusted according to the input.

The input gate operation logic is:

$$\tilde{c}_t = \tanh\left(w_{xc} \cdot x_t + w_{hc} \cdot h_{t-1} + b_c\right) \tag{10}$$

$$i_t = \sigma\left(w_{xi} \cdot x_t + w_{hi} \cdot h_{t-1} + b_i\right) \tag{11}$$

In the equation: $\tanh$ is the hyperbolic tangent activation function, w_{xc} and w_{hc} are the weight coefficients of $\tilde{c}_t$, b_c is the bias term of $\tilde{c}_t$, w_{xi} and w_{hi} are the weight coefficients of the input gate, b_i is the bias vector of the input gate.

The input gate integrates the information from the previous time step and the current time step as new input, selectively retaining it in the current state. Therefore, through the input gate, LSTM can remember important information in the short term and continuously update the current state.

The output gate operation logic is:

$$o_t = \sigma\left(w_{xo} \cdot x_t + w_{ho} \cdot h_{t-1} + b_o\right) \tag{12}$$

In the equation: w_{xo} and w_{ho} are the weight coefficients of the output gate, and b_o is the bias vector of the output gate.

The output gate generates the output at the current moment. The output gate determines the output h_t at the current moment t based on the current moment t input x_t, the hidden layer state h_{t-1} at the previous moment $t-1$, and the latest state c_t.

The state c_t at the current time t is:

$$c_t = f_t \cdot c_{t-1} + i_t \cdot \tilde{c}_t \tag{13}$$

The output h_t at the current time t is:

$$h_t = o_t \cdot \tanh(c_t) \tag{14}$$

σ is the sigmoid activation function, with a value range of $[0,1]$:

$$\sigma = Sigmoid(x) = \frac{1}{1+e^{-x}} \tag{15}$$

The range of values for the $\tanh$ function is $[-1,1]$:

$$\tanh(x) = \frac{e^{x} - e^{-x}}{e^{x} + e^{-x}} \tag{16}$$

The Bidirectional Long Short-Term Memory Network (BiLSTM) is a Long Short-Term Memory Network with forward and backward connections. Bi LSTM utilizes two independent Long Short-Term Memory Network layers, one processing input in chronological order and the other processing input in reverse chronological order, capturing features of the input sequence from both forward and backward directions, making it suitable for processing time-series data.

III. B. 3) Transformer Encoder

Since this paper only deals with recognition and classification tasks, it uses only the Transformer's encoder structure for feature fusion. The most distinctive feature of the Transformer is its attention mechanism.

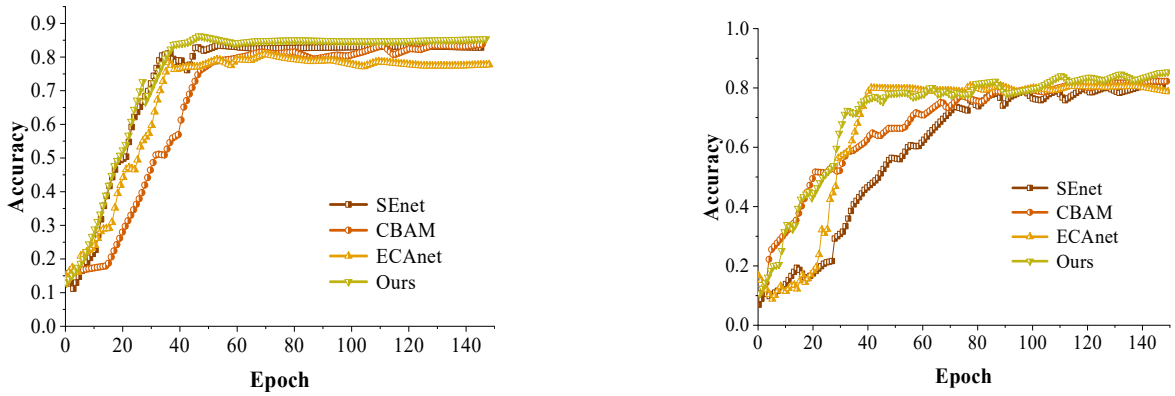
III. C. Experimental Results and Analysis

III. C. 1) Data collection and sources

To validate the effectiveness of the model, it was tested on two datasets. One dataset was CASIA, which included six emotions: anger, fear, happiness, sadness, neutrality, and surprise, with a total of 1,200 voice recordings. The other was Emo-DB, which included seven emotions: anger, sadness, happiness, fear, neutrality, disgust, and boredom, with a total of 534 voice recordings.

III. C. 2) Analysis of experimental results

To validate the effectiveness of the multi-feature fusion-based classroom speech emotion recognition model proposed in this paper, several comparative experiments were conducted on the CASIA and Emo-DB datasets. The results of the comparative experiments between different models are shown in Figure 3. The classroom speech emotion recognition model proposed in this paper achieves the highest accuracy rates compared to SEnet, CBAM, and ECAnet on both publicly available datasets. On the CASIA dataset (Figure (a)), the proposed model converges around 45 iterations, achieving a final accuracy rate of 86.26%. In the Emo-DB dataset (Figure (b)), the proposed model converges after approximately 80 iterations, achieving a final accuracy rate of 85.38%. These comparisons validate the effectiveness of the proposed classroom speech emotion recognition model based on multi-feature fusion.



(a)CASIA dataset (b)Emo-DB dataset

Figure 3: Comparison of experimental results of different models

The comparison of different models on the CASIA dataset for each emotion is shown in Figure 4, and the comparison on the Emo-DB dataset is shown in Figure 5. In the CASIA dataset, the model proposed in this paper achieves higher recognition accuracy than other models for the six emotions: Angry, Fear, Happy, Neutral, Sad, and Surprise. The average recognition accuracy for the six emotions is 0.861, which is 17.17%, 6.63%, and 7.76% higher than the SEnet, CBAM, and ECAnet algorithms, respectively. In the Emo-DB dataset, the average recognition accuracy of the proposed model for the seven speech emotions reached 0.921, outperforming the comparison algorithms by 19.21%, 20.44%, and 16.50%, respectively, thereby validating the effectiveness of the classroom speech emotion recognition model based on multi-feature fusion.

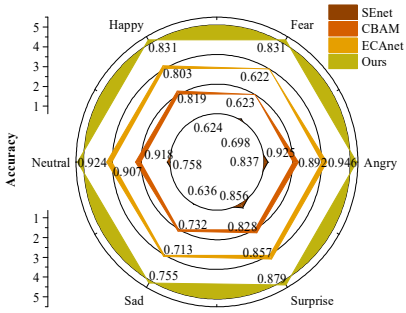


Figure 4: Different models of different emotions in the CASIA data set

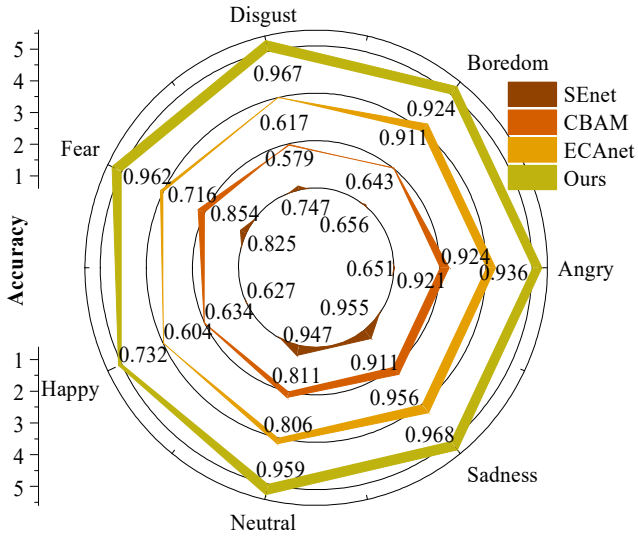


Figure 5: Different models of different emotions in the Emo-DB data set

IV. Practical exploration of intelligent analysis methods for teaching behavior data

IV. A. Construction of Classroom Speech Emotion Dataset

IV. A. 1) Selection of Materials

In order to comprehensively evaluate the proposed intelligent analysis method for teaching behavior data, teaching course materials were obtained from public media platforms. After manual review, classroom videos with relatively clear audio were selected, and finally, 22 classroom materials were manually selected as the data source for this dataset.

IV. A. 2) Data processing

(1) Audio noise reduction

Since the video data was recorded in a real classroom, there may be various types of noise interference in the classroom environment. Therefore, it is necessary to perform noise reduction processing on the original speech segments to extract the purest speech signal possible through preprocessing. This paper uses spectral subtraction for audio noise reduction processing.

(2) Speech Slicing

By using the iFlytek speech recognition API to identify speech segments with timestamps, batch speech data slicing is achieved.

IV. A. 3) Emotional Data Annotation

This paper uses the P value in the PAD three-dimensional emotional space model to represent classroom emotional pleasure, which is used to assess the positive or negative state of an individual's speech emotions. Therefore, we selected a self-annotation format to score the pleasantness of speech P within the range of [-3,3]: the higher the pleasantness, the closer the annotation is to 3 points; the lower the pleasantness, the closer the annotation is to -3 points.

Three members of the project team were invited to annotate the speech data for emotional sentiment, with a pleasantness threshold of [-3,3]. To ensure consistency in the project team members' understanding and recognition of speech emotional categories, 150 speech segments were randomly selected from 1,500 segments for individual annotation. After annotation, the scores from the three annotators were summarized and subjected to consistency testing, yielding a variance of approximately 0.2241, indicating high consistency. Therefore, the three annotators had consistent cognition of classroom speech, and based on this, all classroom audio files were annotated.

IV. B. Interactive Voice Emotion Analysis in Teaching

IV. B. 1) Analysis of emotional change trends

Select a specific classroom level, remove ineffective speech segments such as classroom discussions, exercises, and experiments, and retain six effective interactive speech segments. Analyze these segments using the proposed classroom speech emotion recognition model.

The emotional change curve of the interactive segments is shown in Figure 6. Classroom interactive segment one represents the beginning of the class, primarily consisting of teacher instruction. Before 150 seconds, the teacher's instruction dominates, with the emotional value P primarily ranging from 0 to 1.5, indicating that the teacher's emotions have not yet been fully activated at the start of the class and are in a state of low positive emotion. From 150 to 360 seconds, the overall emotional fluctuations were significant, varying within the range of [-1, 1.25], indicating that during classroom questioning and answering, there were significant fluctuations in emotional assessment values.

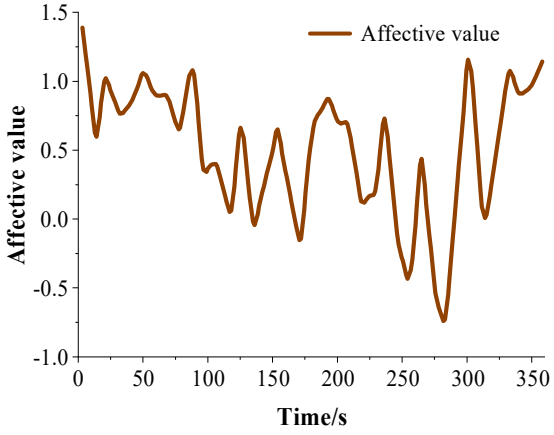
The second classroom interaction segment primarily involves students explaining exam questions, with emotional values P primarily distributed between [1.5, 2.0], indicating a positive and active emotional state, and the classroom is fully activated.

The third interactive voice segment is a practice question-and-answer segment, with overall emotional values P distributed between [1.2, 2.0], indicating that during the question-and-answer phase, the classroom remains active, and students' emotions are positively stimulated.

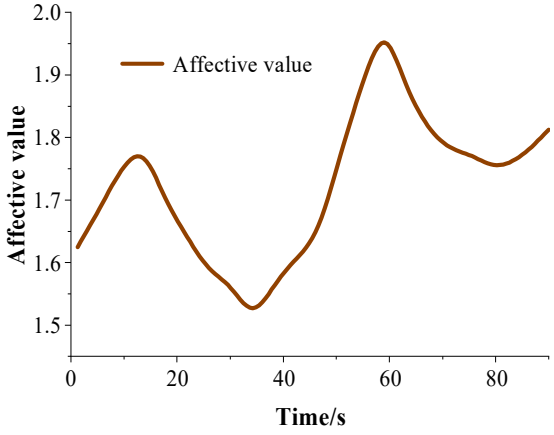
Classroom interaction segment four is a discussion segment, with overall classroom emotional value P distributed within the range [0.85, 1.3]. At 75 seconds, the teacher rewards outstanding students, causing the speech emotional value to rise. At 135 seconds, the teacher asks students a question in a questioning tone, causing the speech pleasantness to drop to around 0.85, indicating the model's effectiveness.

Classroom interaction segment five primarily involves in-class exercise Q&A, with the emotional value P distributed between [0.4, 1.25], showing a significant decrease in pleasant emotions compared to earlier segments, indicating that students' classroom engagement declines in the latter part of the class.

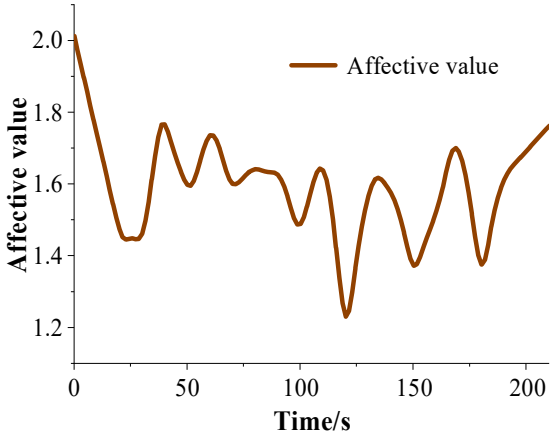
Classroom Interaction Segment Six is the classroom summary section, where the teacher asks students to reflect on their learning gains from the lesson. The classroom emotional tone shows an upward trend, with P values distributed between [0.8, 1.4], indicating that students' emotions are highly elevated during free discussion.

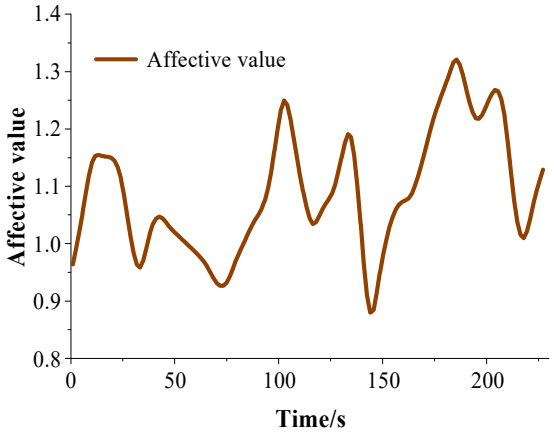


(a)Interaction section 1



(b)Interaction section 2





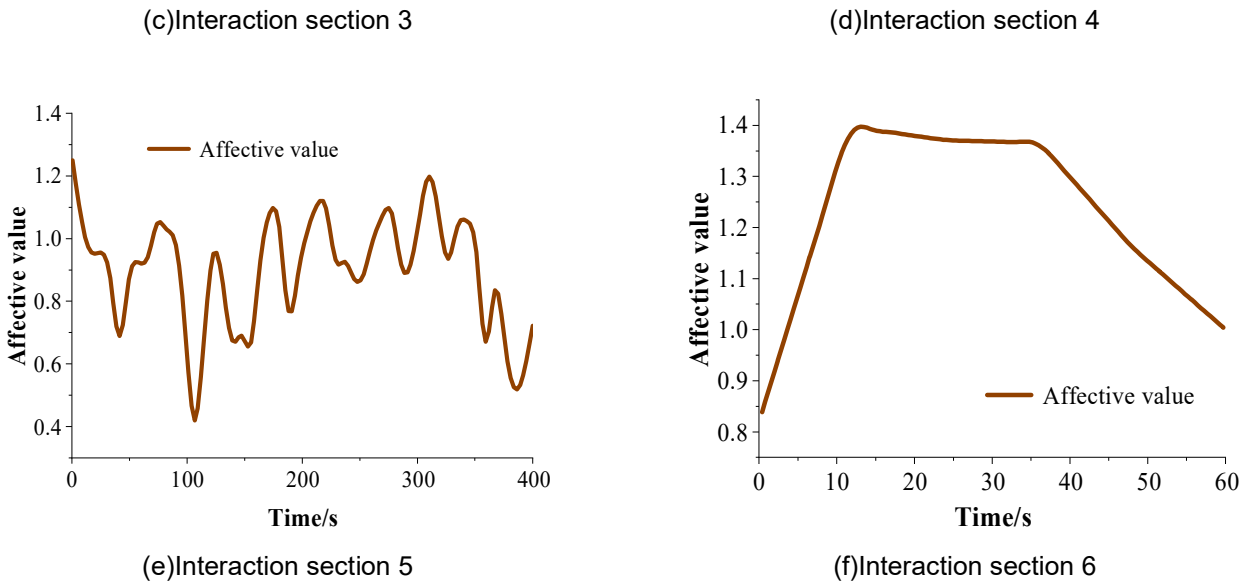


Figure 6: The curve of emotion in the interaction sections

IV. B. 2) Analysis of classrooms with different ratings

To assess the correlation between classroom emotions and classroom evaluations, classrooms with high page views and comments were selected from the national, provincial, and municipal levels of teaching excellence, and a visualization analysis of classroom emotions was conducted for classrooms with different ratings.

The emotional change curve for national-level outstanding classrooms is shown in Figure 7. The overall emotional fluctuation value of the classroom falls within the [0.5, 1.3] threshold range, with relatively small fluctuations in the emotional value P and a generally high overall emotional atmosphere. The emotional change curve for provincial-level outstanding classrooms is shown in Figure 8, with the overall emotional fluctuation range of the classroom falling within [0.2, 0.95], and multiple points of emotional decline observed within the classroom. The emotional change curve for city-level outstanding classrooms is shown in Figure 9. The overall emotional fluctuations in the classroom were within the range of [-1.25, 1], and the overall emotional state of the classroom was relatively low. Combined with the analysis of classroom speech behavior, there was relatively little teacher-student interaction and student participation in the classroom, and the P value was only high when students spoke, indicating that the emotional state of students' speech was slightly higher than that of teachers' speech.

Based on the above analysis, the overall emotional atmosphere of the Ministry-level, Provincial-level, and Municipal-level model classrooms differs, and the emotional pleasure levels are ranked as follows: Ministry-level > Provincial-level > Municipal-level. This indicates that classroom emotions can to some extent influence classroom effectiveness, thereby affecting viewers' evaluations of the classroom.

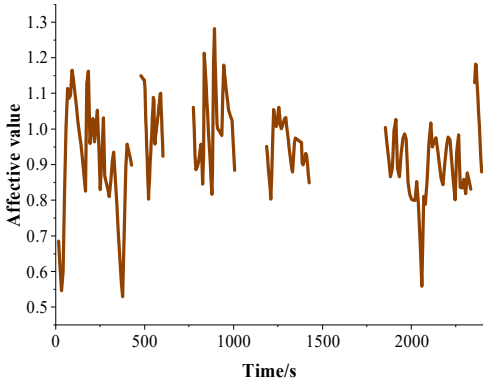


Figure 7: The curve of emotion on ministerial level class

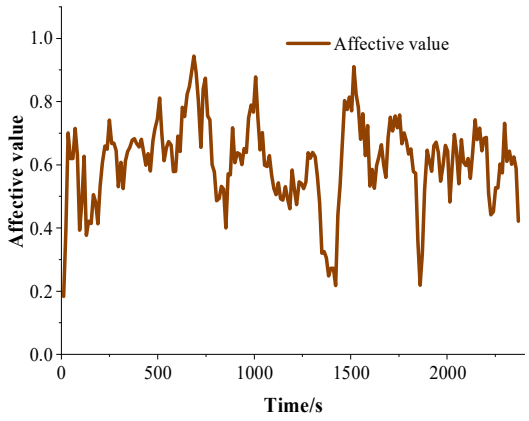


Figure 8: The curve of emotion on provincial level class

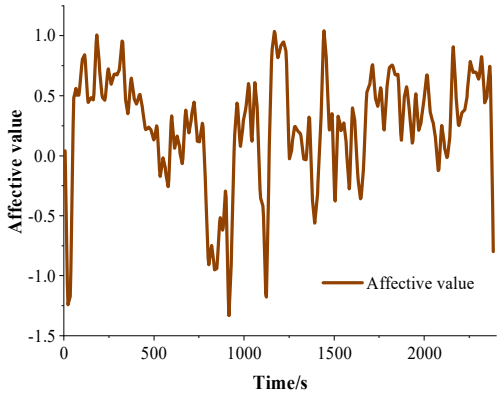


Figure 9: The curve of emotion on urban level class

V. Conclusion

With the continuous upgrading and improvement of intelligent technology, the education industry has also introduced this tool to establish an intelligent teaching environment. This paper uses intelligent models to analyze classroom teaching behavior data and proposes a classroom speech emotion recognition model based on multi-feature fusion. By analyzing speech emotions, it obtains feedback on classroom teaching effectiveness. The performance of this model is analyzed through experiments, and practical explorations of specific teaching behavior analysis are conducted.

(1) The recognition accuracy rate of the constructed classroom speech emotion recognition model exceeds 85% across different datasets, outperforming comparison methods. Additionally, in the recognition of different emotion types, the proposed method achieves recognition accuracy rate improvements of 6.63% to 17.17% on the CASIA dataset and 16.50% to 20.44% on the Emo-DB dataset. This demonstrates the superiority of the proposed method based on multi-feature fusion for classroom speech emotion recognition.

(2) Through emotion recognition of sample classroom interaction segments, the speech emotion values of the six interaction segments were [-1,1.25], [1.5,2.0], [1.2,2.0], [0.85,1.3], [0.4,1.25], and [0.8,1.4], respectively. The emotional trend of classroom interaction shows an initial rise followed by a decline, aligning with the overall progression of the classroom session. The thematic emotional values for national, provincial, and municipal model classrooms are [0.5, 1.3], [0.2, 0.95], and [-1.25, 1], respectively, indicating a certain correlation between classroom emotion and classroom effectiveness.

This paper explores intelligent analysis methods for higher education teaching behavior data and applies speech emotion recognition models in higher education classroom settings. The use of intelligent analysis methods for teaching behavior data can support the harmonious development of higher education teaching, enhance teachers' ability to grasp classroom dynamics, and thereby promote high-quality development in higher education teaching.

Funding

This research was supported by the Jilin Provincial Education Science "14th Five-Year Plan" 2023 General Project: Research on the Logical Basis and Model Selection for Intelligent Analysis of Teaching Behaviors in Universities in the Era of Big Data (Project No.: GH23466), Hosted by Zhang Dan.

Adult Continuing Education Research Project of the Chinese Association of Adult Education "14th Five-Year Plan" 2023 General (Approval No.: 2023-289Y), Title: Research on the Logical Basis and Model Selection for Intelligent Analysis of Teaching Behaviors in Adult Universities under the Background of Big Data, Hosted by Zhang Dan.

References

[1] Wang, Z., Li, L., Zeng, C., Dong, S., & Sun, J. (2025). SLBDetection-Net: Towards closed-set and open-set student learning behavior detection in smart classroom of K-12 education. Expert Systems with Applications, 260, 125392.

[2] Liebowitz, D. D., & Porter, L. (2019). The effect of principal behaviors on student, teacher, and school outcomes: A systematic review and meta-analysis of the empirical literature. Review of Educational Research, 89(5), 785-827.

[3] Korpershoek, H., Harms, T., de Boer, H., van Kuijk, M., & Doolaard, S. (2016). A meta-analysis of the effects of classroom management strategies and classroom management programs on students' academic, behavioral, emotional, and motivational outcomes. Review of educational research, 86(3), 643-680.

[4] Redding, C. (2019). A teacher like me: A review of the effect of student–teacher racial/ethnic matching on teacher perceptions of students and student academic and behavioral outcomes. Review of educational research, 89(4), 499-535.

[5] Gage, N. A., Scott, T., Hirn, R., & MacSuga-Gage, A. S. (2018). The relationship between teachers' implementation of classroom management practices and student behavior in elementary school. Behavioral disorders, 43(2), 302-315.

[6] Madani, R. A. (2019). Analysis of educational quality, a goal of education for all policy. Higher Education Studies, 9(1), 100-109.

[7] Hospel, V., & Galand, B. (2016). Are both classroom autonomy support and structure equally important for students' engagement? A multilevel analysis. Learning and Instruction, 41, 1-10.

[8] Ladd, H. F., & Sorensen, L. C. (2017). Returns to teacher experience: Student achievement and motivation in middle school. Education Finance and Policy, 12(2), 241-279.

[9] Lei, H., Cui, Y., & Zhou, W. (2018). Relationships between student engagement and academic achievement: A meta-analysis. Social Behavior and Personality: an international journal, 46(3), 517-528.

[10] Flower, A., McKenna, J. W., & Haring, C. D. (2017). Behavior and classroom management: Are teacher preparation programs really preparing our teachers?. Preventing School Failure: Alternative Education for Children and Youth, 61(2), 163-169.

[11] Dignath, C., & Büttner, G. (2018). Teachers' direct and indirect promotion of self-regulated learning in primary and secondary school mathematics classes–insights from video-based classroom observations and teacher interviews. Metacognition and Learning, 13, 127-157.

[12] Campbell, S. L., & Ronfeldt, M. (2018). Observational evaluation of teachers: Measuring more than we bargained for?. American Educational Research Journal, 55(6), 1233-1267.

[13] Peng, W. (2017). Research on online learning behavior analysis model in big data environment. Eurasia Journal of Mathematics, Science and Technology Education, 13(8), 5675-5684.

[14] Cantabella, M., Martínez-España, R., Ayuso, B., Yáñez, J. A., & Muñoz, A. (2019). Analysis of student behavior in learning management systems through a Big Data framework. Future Generation Computer Systems, 90, 262-272.

[15] Yan, J., Chen, A., & Wang, H. (2024). An analysis of English learning behavior based on big data and its impact on teaching. Journal of Computational Methods in Science and Engineering, 24(1), 235-251.

[16] Luan, H., Geczy, P., Lai, H., Gobert, J., Yang, S. J., Ogata, H., ... & Tsai, C. C. (2020). Challenges and future directions of big data and artificial intelligence in education. Frontiers in psychology, 11, 580820.

[17] Li, L., Chen, C. P., Wang, L., Liang, K., & Bao, W. (2023). Exploring artificial intelligence in smart education: Real-time classroom behavior analysis with embedded devices. Sustainability, 15(10), 7940.

[18] Jaboob, M., Hazaimeh, M., & Al-Ansi, A. M. (2025). Integration of generative AI techniques and applications in student behavior and cognitive achievement in Arab higher education. International journal of human–computer interaction, 41(1), 353-366.

[19] Hu, J., Huang, Z., Li, J., Xu, L., & Zou, Y. (2024). Real-time classroom behavior analysis for enhanced engineering education: An AI-assisted approach. International Journal of Computational Intelligence Systems, 17(1), 167.

[20] Sainyakit, P., & Santoso, Y. I. (2024). A Classroom Interaction Analysis of Teacher and Students by Using FIACS. Acitya: Journal of Teaching and Education, 6(1), 157-167.

[21] Li, H., & Zeng, X. (2024, June). Analysis of Teacher-Student Interaction in Middle School English Classroom Based on iFIAS. In Proceedings of the 2024 9th International Conference on Distance Education and Learning (pp. 270-277).

[22] Fang, Z., & Sandelius, S. E. (2024, March). Adaptation of Flanders Interaction Assessment System for Online and Offline Comparative Purposes. In 2024 13th International Conference on Educational and Information Technology (ICEIT) (pp. 319-323). IEEE.

[23] Wang, D., Han, H., & Liu, H. (2019, October). Analysis of instructional interaction behaviors based on OOTIAS in smart learning environment. In 2019 Eighth International Conference on Educational Innovation through Technology (EITT) (pp. 147-152). IEEE.

[24] Wang, W. (2024). Application of deep learning algorithm in detecting and analyzing classroom behavior of art teaching. Systems and Soft Computing, 6, 200082.

[25] Shi, S., Gao, J., & Wang, W. (2021). Classroom Teaching Behavior Analysis Based on Artificial Intelligence. Artificial Intelligence in Education and Teaching Assessment, 25-36.

[26] Jasim, A. H., & Hoomod, H. K. (2025, April). Intelligent student behavior recognition system for the classroom environment using hybrid deep learning. In AIP Conference Proceedings (Vol. 3282, No. 1, p. 030011). AIP Publishing LLC.

[27] Jia, Q., & He, J. (2024). Student Behavior Recognition in Classroom Based on Deep Learning. Applied Sciences, 14(17), 7981.

[28] Chen, G., & Zhou, J. (2024, May). Application of Deep Learning in Students' Behavior Analysis and Intervention. In The World Conference on Intelligent and 3D Technologies (pp. 533-543). Singapore: Springer Nature Singapore.

[29] Zou, X. (2024). Design of Student Behavior Analysis and Management System Based on Deep Learning. International Journal of High Speed Electronics and Systems, 2540167.

[30] Gong, B., & Jing, F. (2020, April). Research on the data analysis of college classroom teaching behavior by using deep learning. In International Conference on Education, Economics and Information Management (ICEEIM 2019) (pp. 48-50). Atlantis Press.