

Analysis of Classroom Teaching Behaviors Based on Multimodal Data Model

Xinrui Ma, Jingxia Chen*

Beijing Union University, Beijing, China,

*Corresponding Author.

Abstract: Precise identification of classroom behavior can help teachers and students understand classrooms, and help promote the development of smart classrooms. This article designed a multi-mode data model supported by multimodal data support based on classroom teaching scenes, including classroom teaching layers, data collection layers, algorithm analysis layers, and application service layers. Analysis of classroom teaching behaviors, this article extracts the image characteristics in the video based on the deep learning algorithm of YOLO-V5, the voice recognition technology extracts voice characteristics. Multi-mode data model analysis has obtained good analysis results and recognition results. In order to verify the effectiveness of the selected model, the model performance was performed on the labeled classroom behavior data set. The test results show that the selected model shows good performance in the analysis and identification of classroom behavior in the education scene.

Keywords: Multi-Mode Data; Analysis of Classroom Teaching Behavior; Behavior Recognition; Smart Classroom

1. Introduction

The technology of artificial intelligence has penetrated into every corner of our daily life. The reform of the field of education provides us with technical support and urgent needs. The "New Generation Artificial Intelligence Development Plan" released by the State Council clearly states that needs to make full use of intelligent technology to accelerate the innovation of talent training models and the reform of teaching methods. To this end, a new type of classroom teaching behavior analysis system should be constructed, which covers intelligent learning and interactive learning to

meet the development needs of future education.

Classroom teaching behaviors include teachers' behavior, student behavior and interactive behavior. The study of classroom teaching behavior is to explore the occurrence of teaching behavior and the development law behind the scattered and specific classroom teaching behavior. With the development of technology, the analysis of classroom teaching behavior is also continuously improved and improved, and it has experienced the following stages:

First stage: Statistical analysis stage based on traditional means.

The interactive analysis system (Fias) proposed by Flanders [1] in the 1960s divided the speech of teachers and students in the classroom into teachers' words, student words, stills, or chaos. Artificial coding forms an interactive analysis matrix in the classroom. Gu Xiaoqing [2] and other scholars proposed Information Technology-based Interactive Analysis and encoding System (ITIAS), adding categories of students' speech behavior and technology.

The second stage: informatization stage based on information technology.

Jin Jianfeng et al. [3] Multimedia, interactive electronic whiteboard, and Moodle platforms with three different information technology support analyzed the teaching behavior of teacher-student classroom teaching in information technology curriculum teaching. Li Jing et al. [4] used NVIVO analysis tools to study classroom interaction behaviors of different information teaching environment based on the perspective of qualitative analysis. Mu Su et al. [5] proposed the classroom Teaching Behavior Analysis System (TBAS) and series analysis methods, and built 14 types of classroom teaching behavior variables based on teaching activities. It counts the distribution of teachers and students, interactive behaviors

of teachers and students, and the media in teaching in teaching. Application in the middle. Third stage: intelligent phase based on artificial intelligence.

Sun Zhong et al. [6] based on computer vision technology to classify the teaching scene first, and then identify and calculate the behavior of teachers and students in classrooms, realized the functions of key frame extraction, student tracking, action recognition and action statistics. Liu Xinyun et al. [7] detected 8 student class behaviors such as standing, watching blackboards, recording notes, talking to classmates, and using mobile phones. Lu Guoqing [8] and others automatically marked the activities and behaviors of teachers and students in classroom teachers' videos, and used correlation analysis, non -parameter difference inspection and other methods to analyze the types, laws and differences of classroom teaching behaviors. Polito et al. [9] uses data such as eye movements, brain waves, videos, and voice of teachers with a head -with portable eye movement instrument, brain wave instrument, and smartphone. Activity. Ricklme et al. [10] analyzed the effect of student group collaboration through collecting students' voice data. Chen Yashu [11] analyzed the classroom interaction behavior through multiple dimensions such as classroom emotional level, classroom mode, and classroom interactive structure from multiple dimensions, and identified students' emotional state.

At present, some scholars have applied them to the field of education, but most of them are limited to single -mode analysis, and rarely involve multi -mode integration analysis. In particular, the multi -dimensional data analysis methods and system construction of students' group characteristics in different disciplines. Therefore, it is of great significance to carry out the study of multi -mode teaching behaviors based on this background. Relying on artificial intelligence technology, this study, from the perspective of multi -mode teaching behavior analysis, classified the constituent elements of multi -subject classroom teaching behavior analysis, and designed a multi -model -based classroom teaching behavior analysis model. This model builds a practical framework for classroom teaching behavior analysis, which aims to provide the theoretical foundation for classroom teaching practice supported by multi -mode data, thereby

promoting classroom teaching in the direction of intelligence.

2. Research Method

2.1 Data Collection Principles

In the experiment of this article, we used about 900 minutes of field classroom teaching videos and audio as the source of the original data. The experiment uses behavioral analysis technology. After processing these data, it obtains the main behavior mode and corresponding action indicators in student class learning activities, and uses this as a basis for determining whether it meets the requirements of teaching goals.

2.2 Data Mark

It can be found that teaching behavior can be divided into two categories: teacher behavior and student behavior, which requires multiple target recognition. Among them, the teacher's classroom behavior is marked with textbooks, finger blackboard projection, gesture or two -handed comparison. Students' classroom behaviors are marked as raising hands, operating desktops, walking, writing of paper, looking up, bowing, standing, etc. The selected behavior shows distinctive characteristics, and there should be no ambiguity or duplication between them, which helps better perform behavioral identification (Table 1)

Table 1. The Category and Specific Number of Pictures of the Classroom Behavior Set of Teachers and Students are 50,000.

Teacher Classroom Collection	Student Classroom Behavior Collection
Teaching textbook exercises:113	Make a raised hand gesture:8015
Finger blackboard projection:1458	Lower head and bend down to operate the Desktop:426
NO	Quarter-deck:8115
gestures:7359	Write a board script:11972
Hand gestures:5337	Rise:3108
	Lower your head:1823
	Stand:674
	Raise hands:534
	Lying on the table:960

2.3 Image Acquisition Based on YOLO-v5

Multimodal data models can handle data from different sources and forms, including text, images, audio, video, etc. In the collection of image and video data, this paper uses the YOLO-v5 deep learning algorithm to extract image features from videos. The YOLO-v5

model is decomposed into four parts: input, backbone network, feature fusion network, and output. The input is designed with adaptive image size: regardless of the size of the images in the dataset, the input will automatically scale them to 608 * 608 according to the short edges of the images. If the aspect ratio of the image does not match the input, black will be automatically added to the outer edge of the image, so that the images input to the backbone network conform to 608 * 608 * 3, meeting the consistency of network input. In the backbone network structure, in order to increase the number of layers in the network and prevent gradient vanishing or explosion problems, the CSPNet (Cross Stage Partial Network) model is applied to the ResNet (Residual Network) module. During training and testing, PANet (Path Aggregation Network) is applied in the feature fusion network, and the Spatial Pyramid Pooling (SPP) module is used for feature extraction to improve detection performance. Choosing Leaky ReLU as the activation function has improved the accuracy of the model to a certain extent.

This article proposes a method based on deep feature fusion for video preprocessing using deep learning techniques, combined with convolutional neural networks and attention mechanisms. This article chooses the image correlation coefficient method to extract keyframes and adopts the correlation coefficient algorithm:

$$R = \sum_N \frac{\sum_m \sum_n AB}{\sqrt{\sum_m \sum_n A^2} \sqrt{\sum_m \sum_n B^2}} \quad (1)$$

In the formula, A and B represent any image, m and n are the width and height of the image, respectively. N is the number of channels in the image. After multiple attempts, when the correlation coefficient R is less than 0.8, it can be determined that there is a sudden change in the video, which requires extracting keyframes. In a Bounding box (Bbox) for detection, the trained parameters include the position of Bbox (g), the classification of behavior recognition (α), and the confidence level of target prediction (c). Due to the fact that a good prediction box needs to have the largest overlap area, the closest center point distance, and the closest aspect ratio to the standard box's aspect ratio, for the position of the geometric quantity Bbox, the Euclidean distance is no longer used to measure the

difference between the predicted value and the accurate value. Instead, the intersection and union of the prediction box and the accurate box are used to design its loss:

$$L_g = 1 - (R_0 - \frac{d_1^2}{d_2^2} - \frac{v^2}{(1-R_0+v)}) \quad (2)$$

In the formula, d1 is the geometric distance between the center of the predicted box and the center of the standard box, used to limit the distance between the center points. D2 is the diagonal length of the bounding rectangle between the predicted box and the standard box, used to limit the maximum overlap area. R0 is the ratio of the intersection and union area of the standard box P and the predicted box, calculated as follows:

$$R_o = \frac{S_{(P \cap \hat{P})}}{S_{(P \cup \hat{P})}} \quad (3)$$

V is a parameter used to measure the consistency of aspect ratios, calculated as follows:

$$v = \frac{4}{\pi^2} \left(\arctan \frac{w_{\hat{p}}}{h_{\hat{p}}} - \arctan \frac{w_p}{h_p} \right)^2 \quad (4)$$

In equations WP and W, the width of the standard box and the predicted box are respectively represented, while hp and h represent the height of the standard box and the predicted box, respectively. In the post-processing stage, the selection of prediction boxes adopts non maximum suppression and algorithms, aiming to eliminate the prediction boxes that do not meet the conditions the most. For the behavior category selected by Bbox, since the behavior category is a logical quantity, cross entropy is used for calculation here. Assuming the recognition probability of behavior is p(x), the category loss is:

$$L_\alpha = \sum_{i=0}^N \sum_{j=0}^M -I_{ij}^o (\hat{P}_i(x) \lg(P_i(x)) + (1 - \hat{P}_i(x)) \lg(1 - P_i(x))) \quad (5)$$

In the formula: I is a marker indicating whether the center of the object is in the grid. If it is 1 in the grid, it is not 0. N represents the total number of multi-scale grid pixels at the output end. M represents the number of Bboxes in the grid.

For the prediction confidence c of the target prediction, cross entropy is also used, and the loss introduced by confidence is:

$$L_c = \lambda_c \times \sum_{i=0}^N \sum_{j=0}^M -I_{ij}^o (\hat{C}_i \lg(C_i) + (1 - \hat{C}_i) \lg(1 - C_i)) + \lambda_\alpha \times \sum_{i=0}^N \sum_{j=0}^M -I_{ij}^n (\hat{C}_i \lg(C_i) + (1 - \hat{C}_i) \lg(1 - C_i)) \quad (6)$$

In the formula: λc is the weight containing the target. λn is the weight without the target. The final loss function is the weighted sum of these loss functions. Specifically, since equation (5) has already been weighted, no proportionality coefficient will be added here, that is:

$$L = \lambda_g L_g + L_c + \lambda_\alpha L_\alpha \quad (7)$$

Some key parameters in the formula: $\lambda c=3$, $\lambda n=1$, $\lambda \alpha=2$, $\lambda g=1$.

2.4 Speech Collection Based on Speech Recognition

Convert the collected audio into digital audio signals, perform speech noise reduction processing, and store the processed audio signals in a dedicated database. The preprocessing steps include pre emphasis processing, framing and windowing processing, and endpoint detection processing.

The denoising process is as follows: training pure speech features using PM-DNN (Particle Matter-Deep Neural Network) network [12], calculating and processing, outputting masking threshold T , calculating the correlation between masking threshold and environmental noise amplitude spectrum N , outputting perceptual gain function, merging the calculated G and noisy speech amplitude spectrum Y into a new network layer, and superimposing it with the original network. The final output S is calculated by formula.

$$\hat{S} = G \otimes Y = \frac{1}{1 + \max(\sqrt{\frac{N^2}{T}} - 1, 0)} \otimes Y \quad (8)$$

After calculating the final output S of the network, the weight parameters of the network are trained using pure audio signals. The objective function of training consists of two parts as shown in the formula:

$$J = w \|\hat{S} - S\|_2^2 + v \|\tilde{S} - S\|_2^2 \quad (9)$$

In the formula, w represents the weight of the error between \hat{S} and the pure speech signal S , and v represents the weight of the error between the speech amplitude spectrum S and the pure speech signal S , while ensuring that $w+v=1$.

Pre emphasis in preprocessing is the process of inputting an audio signal into a high pass filter, outputting high-frequency components, and ultimately obtaining an enhanced audio signal. This article uses the overlapping framing method for processing. As shown in the Figure 1, the overlapping framing method ensures a

smooth transition between sub signals by keeping the next frame partially overlapping the previous frame during audio segmentation.

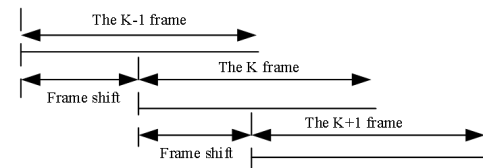


Figure 1. Schematic Diagram of Overlapping Frame Division Method

After completing the frame processing steps of the audio signal, perform windowing operation on the segmented signal. In this article, we choose the window function of Hamming window.

$$w(n) = \begin{cases} 0.54 - 0.46 \cos\left[\frac{2\pi n}{N-1}\right], & 0 \leq n \leq N-1 \\ 0, & \text{others} \end{cases} \quad (10)$$

In endpoint detection, the start and end positions of the audio signal are selected by setting a short-term energy threshold. If $X_n(m)$ is defined as the n th frame audio signal, the formula for calculating the short-term energy of one frame audio signal is as follows:

$$E_n = \sum_{m=0}^{N-1} [x_n(m)]^2 \quad (11)$$

2.5 Classroom Teaching Behavior Analysis Model Based on Multimodal Data Model

The multimodal model is based on the "image sound text" multimodal technology, establishing a layout of multimodal large models and integrating internal images, text, speech, etc. for research and development [13]. It not only has full modal understanding ability, generation ability, and correlation ability, but also can read text, images, and audio, and can combine images, sounds, and videos to complete scene analysis. Its understanding and generation ability is closer to that of humans. Its model is shown in Figure 2.

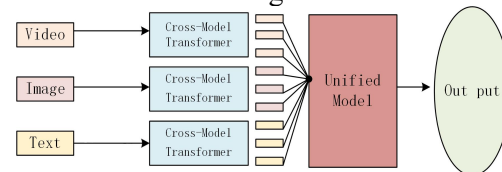


Figure 2. Multimodal Big Data Model

Overall, the analysis process of multimodal learning typically involves data collection, analysis, fusion, and practical applications. In this process, we will demonstrate its effectiveness through different processing methods of the image. In order to provide a clearer description, this study mainly explores

the analysis process of classroom teaching behavior, using multimodal learning methods with a classroom teaching behavior analysis model as the center. Process data plays a guiding role in the data-driven educational background, involving classroom teaching and data collection, as well as in four areas of intelligent analysis and application services: classroom teaching layer, data collection layer, algorithm analysis layer, and application service layer. We have constructed a multimodal data support system and developed a practical framework for classroom teaching behavior analysis (see Figure 3), which provides necessary support for specific classroom teaching activities.

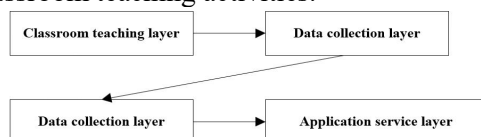


Figure 3. A Practical Framework for Classroom Teaching Behavior Analysis Supported by Multimodal Data

Firstly, the classroom teaching level is based on the scenario design of teaching behavior classification.

This study is based on the classification of various elements of teaching behavior, and the elements of teaching context cover the subject of teaching, the content taught, and the available teaching resources. In terms of teaching environment and teaching interaction:

- 1) The teaching content involved. According to the standards of the curriculum, in order to meet the standards of standards and teaching objectives, teachers need to choose appropriate teaching methods based on the actual situation of daily life. In terms of content, we need to accurately understand the difficulty of knowledge and present it in an appropriate way.
- 2) Equipment used for teaching. As an auxiliary tool for classroom teaching, this tool plays a crucial role in teaching activities such as teacher demonstrations and lectures. For example, by using virtual intelligent devices such as AR and VR, teachers can create a relaxed and realistic learning environment while teaching course content. By learning different scenarios, students' participation in classroom activities can be enhanced.
- 3) Various activities related to teaching. This constitutes the key connection between teachers and students in the classroom, and is also the core of all teaching activities. In order

to achieve this goal, we need to use appropriate teaching tools and methods to ensure that resources match course requirements when designing teaching activities, and integrate key elements such as teaching content, teaching methods, and teaching media throughout the entire activity. This achieves behavioral interaction in the aspects of "multiple subjects, multiple spaces, and multiple tasks". Through mutual decomposition, not only can it help learners master knowledge, but it can also stimulate students' interest in learning, cultivate psychological qualities including but not limited to group cooperation and communication abilities, as well as practical and exploratory abilities. Maximizing the satisfaction of learners' learning needs and expectations.

Secondly, the data collection layer is based on multimodal data for classroom teaching behavior representation.

Analyzing the teaching behavior of multimodal classrooms is actually a study of the behavior and learning methods of teachers. It is a comprehensive description of multiple aspects such as student behavior, emotions between teachers and students, and teacher-student interaction. To fully achieve this goal, it requires a large number of implicit and explicit, multidimensional data indicators. It is displayed from the surface to three dimensions. This article is based on the classification of behavioral elements in multidimensional classroom teaching, and uses intelligent technology, including video, audio, and smart devices, to comprehensively analyze and collect data from various aspects of the teaching process through "multi subject, multi environment, and multi link" methods.

- 1) From the perspectives of learners and teachers: cameras, videos, and intelligent learning tools are used to collect data on learners and teachers regarding their apparent behaviors and potential information.
- 2) In the information space, we adopt the method of knowledge graph to label learning content through surgery and continuously track and record learners' learning progress. Based on interaction data between humans and machines, as well as on-site testing data, personalized and precise resource push operations can be achieved.
- 3) Interaction between teachers and students: Using microphones, microphones, and cameras to capture interactive information between

learners and teachers or peers, including voice information involved in communication, such as teachers' lectures, prompts, and other forms of guidance, learners' voice expressions in answering questions, communicating, and related data on movements and postures. By applying artificial intelligence technology, we can simulate the behavior, language, and facial expressions of teachers and students to conduct in-depth mining and analysis of multidimensional data, and present the results in a visual way. This display method to some extent fills the gap of traditional data collection methods and enhances the ability to analyze classroom teaching behavior comprehensively. Again, the algorithm analysis layer is based on algorithm research using multimodal data.

The algorithm analysis layer is the core of this study, which focuses on using multimodal data for in-depth analysis and research of educational issues. The main process of this layer consists of five steps. 1) Model structure determination. Firstly, it is necessary to clarify the network structure of the large model, including the types, quantities, and connection methods of each layer. Common neural network structures such as Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN), and Transformers have their unique advantages and applicable scenarios. The algorithm used in this study is based on YOLO-v5 deep learning algorithm to extract image features from videos. Speech recognition technology extracts speech features. 2) Calculation of parameter quantity. Calculating the number of parameters needs to be done layer by layer, including weights and biases. Weight is used to adjust the weight of input data. Bias is used to adjust the output of the model. By summarizing the number of parameters in each layer, the total number of parameters in the entire model can be obtained, providing data support for subsequent applications. 3) Algorithm execution efficiency evaluation. Including the time and space complexity of the algorithm, evaluate the execution efficiency of the algorithm, and optimize the performance of the model accordingly. 4) Algorithm optimization and adjustment. Optimize and adjust the algorithm according to the actual situation. This includes selecting appropriate optimization algorithms (such as gradient descent and its variants), introducing regularization techniques to

prevent overfitting, and utilizing self attention mechanisms to improve model performance. In addition, a strategy of pre training and fine-tuning is adopted to improve the generalization ability of the model by pre training on large-scale data, and then fine-tuning on specific tasks to enhance performance. 5) Visualization and interpretation of results. The algorithm analysis process presents the results in an intuitive and understandable way, including the use of visualization techniques such as charts and animations to showcase the performance, parameter distribution, and other information of the model. At the same time, it is necessary to explain and clarify the analysis results, understand the working principle and performance of the model.

Finally, at the application service layer, we adopt an intelligent teaching method based on multimodal data.

The analysis of teaching behavior in multimodal classrooms is mainly in the stage of application services. We need to study how to truly integrate the analysis data from classroom teaching into teaching, and ensure that the value of classroom teaching data reaches its optimal state in all aspects of teaching and management. 1) Personalized chemistry is based on teaching behavior data from multimodal classrooms, providing learning services. The analysis and research of multimodal classroom teaching behavior is conducted through data collection and analysis. By analyzing learners' multimodal data, we can better understand their cognition, behavior, and emotions. Accurate evaluation helps strengthen the connection between learners and educational environment elements. By establishing close connections, we can clarify how learning occurs and adjust teaching methods based on situational factors, adjust the direction of learning, and provide learners with customized support for learning with certainty. 2) The service provided by utilizing teaching behavior data from multimodal classrooms for efficient teaching methods, by collecting and analyzing the learning process and accompanying patterns of learners, teachers can gain a deeper understanding of students' learning habits and cognitive abilities. Furthermore, we need to adjust our teaching methods and transform traditional teaching methods to enhance the teaching effectiveness of the classroom. In terms of teaching behavior

data from multimodal classrooms, it also includes classroom records automatically generated from calculation results based on these data, as well as the growth trajectory of teachers. According to the research of Liu Qingtang et al. in 2019, it is necessary to monitor the changes in teachers' teaching abilities in a timely manner. Teachers' self reflection can provide important references for improving their professional teaching quality. 3) To provide services for classroom management using data from multimodal classroom teaching behaviors and utilizing intelligent devices for "multi space, multi subject, and multi link" data collection, we need to build a unified classroom activity arrangement. This model can to some extent meet the dynamic structured requirements of data-driven precise teaching needs and help improve the management efficiency of classroom teaching. Teachers should be good at identifying problems in students' thinking and guiding them in a timely manner, helping them form new knowledge systems and methods and strategies. At the same time, adopting data-driven multi cognitive tracking and attribution analysis methods can comprehensively evaluate learners' cognition, thinking, skills, and emotions, assist teachers in accurately grasping students' learning status and current learning status, provide guidance for targeted teaching, and improve the service quality of classroom teaching management.

3. Results and Analysis

3.1 There is a Correlation between Teacher Behavior and Student Behavior

The correlation between teacher behavior and student behavior is crucial for maintaining teaching order and improving teaching effectiveness. The results of multimodal data model analysis show that teachers' patrol behavior significantly affects students' classroom performance. 1. There is a correlation between teachers' inspection behavior and students' focused emotions. Teachers' inspection behavior can induce students to develop focused emotions and reduce their chaotic behavior. There is a significant positive correlation between teachers' inspection behavior and students' classroom participation. The teacher's eye contact, body movements, and other diverse

information effectively enhance students' classroom attention and participation. Specifically, the teacher's inspection behavior allows students to feel more attention and encouragement, thereby actively participating in classroom discussions and activities. At the same time, teacher inspections also help teachers to promptly identify students' confusion and problems, adjust teaching strategies in a timely manner, and ensure teaching effectiveness. This positive cycle mechanism further enhances students' classroom participation and promotes their learning and development. 3. Teachers' teaching behavior has a direct impact on students' classroom quietness. The analysis results of the multimodal data model show that clear and interesting lectures can attract students' attention, reduce their distraction and chaotic behavior. On the contrary, if the teacher's lectures are dull or difficult to understand, it can lead to chaotic classroom order.

3.2 The Proportion of Teacher Behavior is Higher than That of Student Behavior

Through multimodal data model analysis, it is shown that the proportion of teacher behavior is as high as 78%, which is significantly different from the proportion of student behavior. The Rt value of teacher behavior occupancy rate is as high as 78%, indicating that teacher behavior occupies most of the classroom time. So the proportion of student behavior (S) is relatively low, indicating that students' activity time, thinking time, discussion time, etc. in the classroom may be compressed, which is not conducive to students' active participation and deep learning. The behavior conversion rate Ch value represents the ratio of the number of mutual conversions between teacher behavior (T) and student behavior (S) during the teaching process to the total number of behavior samples. In this example, due to the high proportion of teacher behavior, it may lead to a low rate of behavior conversion, which means that there are relatively few opportunities for classroom interaction, student feedback, and so on.

3.3 The Frequency of Teachers' Creative Questioning is Positively Correlated with the Frequency of Students' Positive

Classroom Behavior

The results of the multimodal data model show that teachers' creative questioning, especially open-ended and complex questions, is positively correlated with the frequency of student behaviors such as answering, raising hands, and group discussions. Creative questioning promotes teaching interaction, making communication and feedback between teachers and students, as well as between students, more frequent and effective. Openness and complexity issues can stimulate students' learning motivation and interest, making them more actively participate in classroom activities and improve learning outcomes. The increase in the frequency of student behavior reflects their level of learning engagement, which helps cultivate their ability for self-directed learning and innovative thinking. The increase in the frequency of student behavior reflects their level of learning engagement, which helps cultivate their ability for self-directed learning and innovative thinking.

4. Conclusion and Reflection

Overall, artificial intelligence technology provides a new perspective and conceptual thinking for classroom teaching, innovative evaluation methods, and situations that are difficult to describe and visually perceive in a data-driven manner. This may prompt teachers to take action in teaching decision-making and classroom management, and we are moving towards a more scientific and efficient direction. Multimodal data models utilize deep learning techniques to preprocess data from different sources (such as text, images, etc.) and convert them into numerical vectors for model processing. Based on the analysis of multimodal data models, it is shown that there is a correlation between teaching behavior and student behavior. In order to improve the quality of teaching, teachers should pay attention to the use of patrol behavior and optimize teaching strategies to create a more conducive classroom atmosphere for students' learning. 2. The proportion of teacher behavior is higher than that of student behavior. Teachers should pay attention to balancing teacher behavior and student behavior, improving classroom interaction and student participation, and emphasizing the balance between teaching leadership and student

subjectivity. 3. The frequency of creative questioning by teachers is positively correlated with the frequency of positive classroom behavior by students. Creative questioning requires teachers to transform from traditional knowledge transmitters to guides and partners for students. Teachers need to constantly learn and explore new teaching methods and strategies to meet the needs of students. This transformation is of great significance for improving teachers' professional competence and teaching quality.

The purpose of this study is to explore the core factors and principles of multimodal classroom teaching behavior analysis based on the current research status. A model for analyzing classroom teaching behavior is constructed, and practical analysis of classroom teaching behavior is proposed. This provides a framework for future classroom teaching and solid theoretical support and reference for practical analysis and research.

Acknowledgments

This work is supported by the R&D Program of Beijing Municipal Education Commission(KM202111417002).

References

- [1] Flanders N A. Intent, action and feedback: A preparation for teaching. *Journal of Teacher Education*, 1963, 3(14): 25-260.
- [2] Xiaoqing Gu, Wei Wang. New exploration of classroom analysis technology to support teachers' professional development. *China Electronic Education*, 2004(7): 18-21.
- [3] Jianfeng Jin, Xiaoqing Gu. Analysis and Research on Classroom Teaching Behavior in the Information Technology Environment. *China Electronic Education*, 2010(9): 82-86.
- [4] Jing Li, Qi Zhang, Zhigang Miao, etc. Research on Interactive Behavior of Information Classroom Teaching in Middle Schools: A Qualitative Analysis Perspective. *China Electronic Education*, 2014(2): 101-107.
- [5] Su Mu, Pingping Zuo. Research on the Analysis Method of Classroom Teaching Behavior in the Information Technology Teaching Environment. *Research on Electronic Education*, 2015, 36(9): 62.
- [6] Zhong Sun, Kaiyue Lv, Luo Liming, etc.

- Analysis of Classroom Teaching Based on Artificial Intelligence. *China Electronic Education*, 2020(10): 15-23.
- [7] Xinyun Liu, Shiping Ye, Denghui Zhang. Improved multi-objective regression method for detecting student classroom behavior. *Computer Engineering and Design*, 2020.
- [8] Guoqing Lu, Kui Xie, Qingtang Liu, etc. Analysis of Classroom Teaching Behavior Based on Artificial Intelligence Engine Automatic Annotation. *Open Education Research*, 2021.
- [9] Prieto L P, Sharma K, Dillenbourg P, et al. Teaching analytics: towards automatic extraction of orchestration graphs using wearable sensors. *Proceedings of the sixth international conference on learning analytics & knowledge*, 2016: 148-157.
- [10] Riquelme F, Munoz R, Mac Lean R, et al. Using multimodal learning analytics to study collaboration on discussion groups: A social network approach. *Universal Access in the Information Society*, 2019, 18: 633-643.
- [11] Lele Zhang, Xiaoqing Gu. Classroom teaching behavior analysis model and practical framework supported by multimodal data. *Open Education Research*, 2022, 28(06): 101-110.13966/j.cnki.kfjyyj.2022.06.011.
- [12] Yashu Chen. Multidimensional Classroom Interaction Analysis Based on Speech Recognition. *Huazhong Normal University*, 2021.
- [13] Zhihui Zhao, Yi Zhou, Weihong Li, Chaohui Tang, Qiang Guo, Rigao Chen. Construction of TCM syndrome differentiation model of type 2 diabetes based on deep learning multimodal fusion. *World Science and Technology - Modernization of Traditional Chinese Medicine*, 2024, 1-1.