Advances in Computer and Communication, 2025, 6(1), 20-26
https://www.hillpublisher.com/journals/acc/
ISSN Online: 2767-2875

Construction and Empirical Study of Intelligent Recognition and Analysis Model of Multimodal Classroom Behavior

Tong Su

Shanghai Lixin University of Accounting and Finance, Shanghai 201209, China.

How to cite this paper: Tong Su. (2025) Construction and Empirical Study of Intelligent Recognition and Analysis Model of Multimodal Classroom Behavior. *Advances in Computer and Communication*, 6(1), 20-26.
DOI: 10.26855/acc.2025.01.004

Received: December 29, 2024
Accepted: January 27, 2025
Published: February 25, 2025

[*]**Corresponding author:** Tong Su, Shanghai Lixin University of Accounting and Finance, Shanghai 201209, China.

Abstract

Under the empowerment of emerging technologies, smart classrooms provide rich educational multimodal data for monitoring, evaluating, providing feedback, and warning students about learning behaviors. Fully collecting, processing, and analyzing educational multimodal data can serve as a reference for educational research and practice. Firstly, by reviewing existing studies on smart classroom learning behavior analysis and multimodal data teaching research, drawing on the learning behavior classification system, a three-dimensional multimodal data analysis framework based on multimodal data types—audio data, image data, and text data—is constructed to present changes in students learning behaviors from four aspects: verbal learning activities, positional movements, bodily actions, and technology usage. Secondly, multimodal data representing students' learning behavior characteristics are encoded and qualitatively characterized to form an encoding system. Lastly, from the perspective of situational and temporal behavior analysis, frequency changes and periodic changes in learning behavior multimodal data from lesson examples are analyzed. The study concludes that under the smart classroom environment, students' participation, initiative, and focus in class learning are continuously improving, the classroom learning atmosphere is developing positively, and ineffective and irrelevant learning behaviors are gradually decreasing. Additionally, by analyzing the multimodal data of students learning behaviors, it aims to help students understand their own learning behaviors and states, and enable teachers to conduct personalized teaching. In addition, researchers construct a learning evaluation system to provide data sources and scientific basis.

Keywords

Smart classroom; learning behavior; multimodal data; framework construction; data analysis

1. Introduction

In today's education sector, a new round of "classroom revolution" is taking place, with the application of technologies such as the Internet of Things (IoT), big data, cloud computing, and artificial intelligence in educational settings, which has brought about a disruptive impact on traditional education while also presenting the most direct challenges. In 2008, IBM proposed the "Smart Planet" strategy, where the concept of "smart" gained widespread attention in education and teaching. The country places great emphasis on the development and application of smart classrooms: it proposed the "cloud-end" information platform architecture based on Smart Classroom 1.0, released the "Ten-Year Development Plan for Educational Informatization (2011-2020)" to promote the development of smart classrooms, and in 2019, the Ministry of Education announced the creation projects of "Smart Education Demonstration Zones" to facilitate the comprehensive application of smart classrooms [1-3].

2. Multimodal analysis of students' learning behaviors in smart classroom

2.1 Research on learning behavior in smart classroom

Teaching is an activity composed of the teachers' "teaching" and the students' "learning", representing the occurrence of a behavior. Classroom teaching behaviors are categorized into teacher behaviors, student behaviors, and interactive behaviors. Among these, student behaviors refer to the behavioral feedback of students on classroom activities during the teaching process. A detailed analysis of behaviors helps teachers understand students learning states and provide more accurate assistance [4, 5]. Classroom student behavior analysis systematically collects all information related to classroom student behaviors, evaluates them according to training objectives and requirements, and uses reasonable evaluation methods to measure and value judgments of their learning activities and related factors.

Multimodal learning analysis is grounded in practical needs integrating learning analysis theory and artificial intelligence technology utilizing intelligent sensing devices and intelligent analysis techniques to explore the full picture of the learning ecosystem and deepen learning analysis. This study attempts to integrate multimodal data from the perspective of multimodal learning analysis to research the characteristics and changes in learning behaviors in smart classrooms [6, 7].

2.2 Analysis framework of multimodal data

This study, based on the characteristics of data types presented in the smart classroom environment, divides students learning behaviors into multimodal data including Sound data (student classroom activities primarily involving speech), Image data (learning activities mainly involving body movements and positional changes), and Text data (learning activities primarily involving technology use). The framework for multimodal analysis of students' learning behaviors in smart classrooms is constructed from three dimensions: sound, image, and text [8]. According to the author's actual observations of classroom learning activities, multimodal data under classroom learning activities are categorized into "invalid" and "valid" data from the third gradient, ultimately generating the multimodal data framework as shown in Figure 1.

2.3 Interpretation of multimodal analysis framework

2.3.1 Multimodal sound data

The smart classroom adopts speech recognition technology, utilizing microphones and audio pickup devices (talking mics) to collect learners' voice data. Language information is the primary carrier of learning activities and the occurrence of learning, and the analysis of learning behaviors should start from the students' verbal interactions in the classroom. Based on the "spoken" and "silent" characteristics of students' vocal data, multi-modal voice data is divided into "spoken data" and "silent data," and further categorized into "effective" and "ineffective" in the third dimension. The author refers to sounds that aid learning as "effective spoken data," while those that do not are termed "ineffective spoken data." "Effective spoken data" includes answering questions, asking questions, sharing results, discussing and communicating, evaluating and providing feedback; "ineffective spoken data" includes topic deviation, idle chatter, etc. The author also refers to silent data that aids learning as "effective silent data," such as thinking, listening, etc.; conversely, it is termed "ineffective silent data," such as zoning out, silence, etc. [9].

2.3.2 Multimodal text data

Learning activities primarily focused on technology usage mainly rely on various online learning platforms (WeChat, QQ, Chaoxing Learning Pass, China University MOOCs, and Baidu Cloud) to achieve human-computer interaction. Cloud learning platforms can generate a large amount of personalized text data, such as data acquisition from smart devices, mobile sign-ins, platform responses, communication discussions, homework exercises, outcome presentations, and evaluation feedback. Based on whether this behavior aids learning, the author categorizes it into "effective technical support" and "ineffective technical support" from the third level. These technical supports can generate personalized learning text records on the learning platform, forming the multi-modal text data source for this study [10].

2.3.3 Multimodal image data

Through video acquisition utilizing motion sensing systems, motion capture systems, and 360-degree video technology, the data of students' positional movements, individual actions, facial expressions, and other image data is digitized for representation, thereby studying the multimodal image data of students' learning behaviors. For example, in positional movement data, researchers study from a spatial perspective, using machine learning algorithms to reveal the impact of learners' and learning groups' positions and distances on their learning states, understanding the students' learning activities. Body action data mainly collects data on students' eye fixation directions and torso orientation movements [11].

3. Research design

3.1 Research subjects

This study selects two sessions, one at the beginning of the semester, one in the middle of the semester, and one at the end of the semester, totaling six sessions of smart classroom teaching activities as analytical cases, observing and analyzing students classroom learning behaviors, adhering to the multi-modal analysis framework of student learning behaviors in smart classrooms, and using multi-modal data analysis methods for smart classrooms to explore the characteristics and patterns of students learning behaviors in smart classroom environments. The selection of cases follows four principles: (1) consistency in classroom types to minimize the impact of class type differences on learning behaviors; (2) clear audio recording to accurately identify students voice information; (3) multi-angle video recording including panoramic views, close-ups, and dynamic captures to observe students overall and partial behaviors; and (4) stable network learning platforms to ensure the smooth operation of smart classroom teaching activities.

3.2 Data encoding

Refine the learning behavior analysis framework and list the corresponding learning behavior characteristics. At the same time, encode and label according to the listed learning behavior characteristics. The sound data includes 10 learning behavior characteristics, forming S1-S10 codes sequentially based on the English initials of the sound data; the position movement in image data includes 6 learning behavior characteristics, forming M1-M6 codes sequentially in the same manner; body movements include 13 learning behavior characteristics, forming A1-A13 codes sequentially; text data includes 10 learning behavior characteristics, forming T1-T10 codes sequentially.

3.3 Qualitative representation of learning behavior

This study will qualitatively characterize the multimodal data reflecting students' learning behaviors. From the external manifestations and internal motivations of learning behaviors, learning behaviors are categorized into participatory learning behaviors, proactive learning behaviors, and focused learning behaviors. Participatory learning behaviors refer to students' actions in completing learning activities in class, representing their "doing" behavior. Proactive learning behaviors reflect students' conscious and proactive learning behaviors, representing their "attitude" behavior. Focused learning behaviors reflect students' mental states in class, representing their "ability" behavior. Among the three qualitative characterizations of learning behaviors, since proactive learning includes participation and participation only reflects students' classroom participation behavior, in this study, if a particular behavior has a proactive character, it is categorized as proactive in data statistics and not counted as participatory for repeated calculations.

3.4 Multimodal data statistics

Utilizing the sound, image, and text data collection devices in the smart classroom environment, two observers observed and statistically analyzed the multi-modal learning behaviors in real time, encoding them with time intervals for comparison. They independently encoded the lessons at intervals of 20 seconds using ELAN6.1 software and finally visualized the statistical results. Based on the learning behavior coding system, the two observers agreed on specific rules and details for data collection, unified them, and conducted data statistics according to the coding system table, organizing and verifying the coded data to provide precise multi-modal data support for learning behavior research.

4. Multimodal data analysis

Behavior frequency analysis is one of the methods for studying learning behaviors referring to the ratio of the number of samples of a particular behavior to the total number of all behaviors in the teaching process. The usage frequency of behavioral categories can reflect the varying frequencies of different learning behaviors among students in the teaching process thereby determining the patterns of changes in students' classroom learning behaviors. The following research conclusions present and analyze multimodal data from the perspective of smart classroom scenarios.

4.1 Participatory learning behavior

Participatory learning behaviors are primarily manifested in three modal dimensions of student position movement (M), bodily actions (A), and technical support for textual data (T), representing students "doing" behaviors. Observers and researchers discussed and decided to collect 17 coded data points across these three dimensions, totaling 65 groups of effective position movement modal data (M1-M4) and 281 groups of effective bodily action modal data (A6-A10) from students in smart classrooms. Effective technical support textual data (T1-T8) were derived from the student learning platform, resulting in a total of 130 data groups. Due to the inability to obtain ineffective behaviors and no-behavior data from existing smart classroom environment equipment, and the subjective differences in human statistics leading to collection difficulties, these were not included in this study. Therefore, the focus of participatory learning behavior analysis is on effective participation behaviors. Word cloud diagrams are formed based on the behavioral coding data from these three dimensions for visual presentation, as shown in Figure 1.

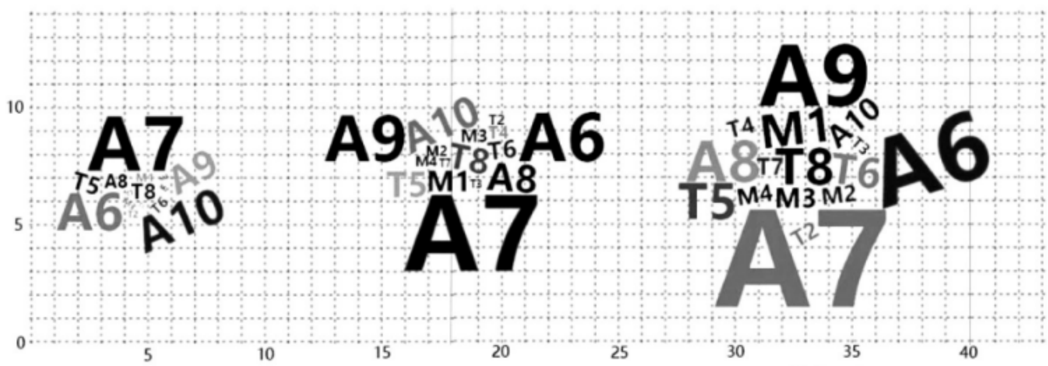

Figure 1. Multimodal data analysis of participatory learning behavior.

4.2 Proactive learning behavior

According to the qualitative characteristics of learning behaviors in the multimodal coding table of student learning behaviors, proactive learning behaviors are mainly manifested in language-based teaching activities, reflecting students' "attitude" towards learning, i.e., multimodal audio data (S). Statistics on lesson examples at three-time points—beginning of the semester, middle of the semester, and end of the semester—yielded a total of 124 sound data points for proactive learning behaviors (S1, S3, S4, S5, S6), including 36 at the beginning of semester, 38 in the middle of semester, and 50 at the end of semester. Due to the inability to accurately obtain behavioral data for S8 and S9, which were ambiguous based on human judgment, they were not included in the statistics. Consequently, the actual proactive learning behavior data generated by students amounted to 84 points, including 16 at the beginning of the semester, 26 in the middle of the semester, and 42 at the end of the semester. Frequency analysis was performed on the five sets of data, yielding the ratio of the actual proactive learning behaviors generated by students in a class to the total number of proactive behaviors expected in that class. A ratio closer to 1 indicates a higher frequency of proactive learning behaviors. The frequency changes show that the ratios of the five sets of coded data increased over the semester and continuously approached 1.

Through the analysis of students' proactive learning behavior frequency, it can be observed that the proactive learning behaviors of students which occur less frequently at the beginning of the semester in smart classrooms gradually increase over time, effectively improving their learning attitudes. For example, students' proactive learning behaviors such as raising their hands, posting achievements, and evaluating online using learning tablets, as well as actively answering questions and engaging in discussions offline, have increased.

4.3 Focus on learning behavior

With the development of the semester, the number of times teachers remind students about their focus on learning content decreases, and students can timely concentrate on the interactive whiteboard (A1), listen attentively to the teachers' lecture (A2), participate in class discussions (A3), and watch the textbook content (A5), with the focus ratio gradually approaching 1. Research findings indicate that at the beginning of the semester, the ratio of students' gaze to learning aids (A4) exceeds 1, indicating that students are overusing learning aids and shifting their attention to the aids themselves. As the semester progresses, this behavior frequency continuously decreases and approaches 1, suggesting that the use of learning aids for non-learning purposes is gradually decreasing, and students' focus on class is gradually increasing.

In teaching research, time-series analysis typically refers to representing the changes in behavior over time in the classroom process using the time axis on the horizontal axis and the behavioral categories on the vertical axis, achieving the behavioral analysis of complex teaching systems. To analyze the dynamic changes in student learning behaviors in smart classrooms from a time-series perspective, the author selected three teaching activity segments—discussion and exchange, outcome sharing, and evaluation feedback—and analyzed them separately. The multi-modal behavioral time-series diagrams of student learning in smart classrooms at the beginning, middle, and end of the semester were plotted (with time series on the horizontal axis and multi-modal co-occurrence on the vertical axis), as shown in Figure 2. Through comparison, it is evident that from a time-series perspective, student learning behaviors exhibit the following more pronounced characteristics.

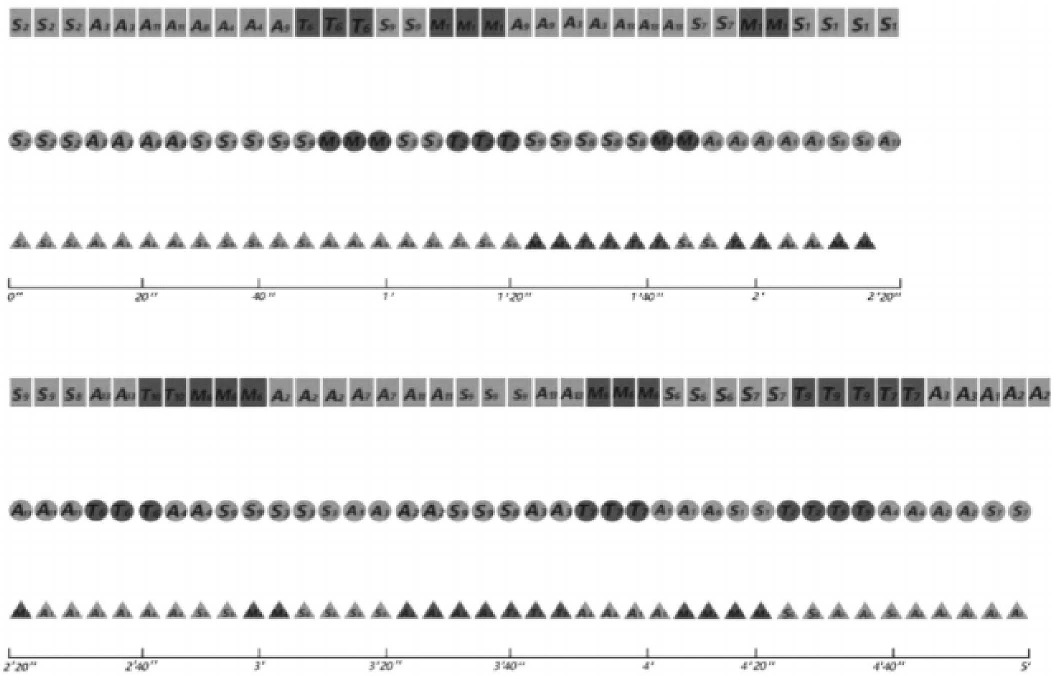

Figure 2. Time sequence diagram of multi-modal behavior of students learning in the smart classroom (1).

Classroom communication and discussion is a process where students participate in learning, engage in autonomous learning, and construct knowledge meaningfully. In the communication and discussion phase of a smart classroom, changes in student behavior are as follows: teachers post discussions on the learning platform, and students collect information and engage in autonomous learning based on discussion questions using smart tools (T8, T4). The elements involved in changes in students' physical movements are sequentially interactive whiteboards (A1, A6), textbooks (A5, A10), learning aids (A4, A9), and peers (A3, A8). The frequency of movement within and between student groups (M3, M4) increases. As the semester progresses, physical movements unrelated to learning (A11), no physical movements (A12, A13), and technical usage behaviors unrelated to learning (T9) among students significantly decrease.

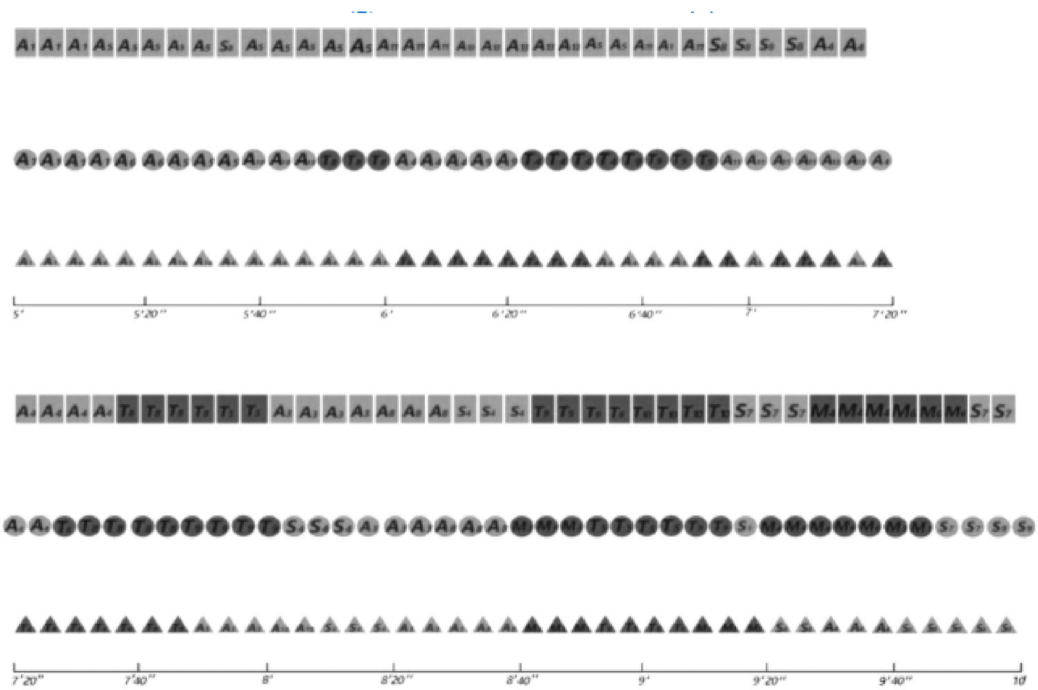

Figure 3. Time sequence diagram of multi-modal behaviors of students in the smart classroom (2).

5. Conclusion

The S-I-T Smart Classroom Student Learning Behavior Multimodal Analysis Framework fully utilizes intelligent devices in the smart classroom environment to collect and generate audio, video, and text data. Based on the multi-modal data types, it forms the S-M-A-T Smart Classroom Student Learning Behavior Multimodal Analysis Coding System, which can effectively present changes in student learning behaviors under the smart classroom from four dimensions: verbal learning activities, positional movements, bodily actions, and technology usage, and endows these behaviors with qualitative representations, providing data collection and behavioral evaluation standards for multi-modal student behavior research.

multimodal data learning and analysis methods are more suitable for studying students learning behaviors in smart classrooms from single-modal effective audio data S analysis and single-modal effective image data A analysis to multi-modal A-T-M tri-modal cross-analysis and then to S-M-A-T four-modal temporal analysis. There are a large number of rich and diverse student learning behaviors in smart classrooms. As the semester progresses, students learning behaviors show significant changes. Analyzing the frequency and patterns of student learning behaviors from both situational and temporal perspectives can provide new ideas for multimodal learning behavior research.

References

[1] Álvarez IM, Manero B, Romero-Hernández A, Cárdenas M, Masó I. Virtual reality platform for teacher training on classroom climate management: Evaluating user acceptance. Virtual Reality. 2024.

[2] Moore TC, Daniels S, Taylor KLH, Oliver RM, Chow J, Wehby JH. Supporting teachers' effective classroom and behavior management: What do teachers tell us? Preventing School Failure: Alternative Education for Children and Youth. 2024.

[3] Herman KC, Reinke WM, Dong N, Bradshaw CP. Can effective classroom behavior management increase student achievement in middle school? Findings from a group randomized trial. Journal of Educational Psychology. 2022.

[4] Lohmann MJ, Randolph KM, Oh JH. Classroom management strategies for Hyflex instruction: Setting students up for success in the hybrid environment. Early Childhood Education Journal. 2021.

[5] Chen T, Mei X. Empirical study on user portrait identification and recommendation of multi-source data fusion. Information Theory and Practice. 2024;47(04):171-180. doi:10.16353/j.cnki.1000-7490.2024.04.022.

[6] Fang H, Hong X, Shu L, et al. Research on teachers' teaching ability analysis framework and its application based on classroom

intelligence analysis model. Modern Educational Technology. 2024;34(02):43-52. doi:10.3969/j.issn.1009-8097.2024.02.005.

[7] Gao D, Wu H, Li Z. Construction and empirical research on patent information. Scientific and Technological Intelligence Research. 2024;6(03):69-82. doi:10.19809/j.cnki.kjqbyj.2024.03.006.

[8] Shenzhen University. A behavior recognition system and recognition method for a multi-mode sensor: CN110807471B [Patent]. 2024-02-02 [cited 2024-12-21]. Available from: https://www.cqvip.com/doc/patent/3338840999.

[9] Shen Y, Huang W. Research and application of grid governance work order generation based on multimodal large model. Jiangsu Communications. 2024;40(01):87-91+96.

[10] Yin B, Wang X, Sun X, et al. Multimodal decoding of teacher classroom management behavior: Behavioral characteristics, classification identification, and timing development. Electrochemical Education Research. 2024;45(10):101-109. doi:10.13811/j.cnki.eer.2024.10.014.

[11] Tang Q, Zhang H, Wu Y. Construction of a multimodal learning analysis model in the intelligent classroom. Journal of Teacher Education. 2024;11(05):49-58. doi:10.13718/j.cnki.jsjy.2024.05.006.