Research on Student Behavior Recognition and its Application based on Machine Learning

Lei Liu

Department of Guangdong Open University (Guangdong Polytechnic Institute)
Guangdong Open University, Guangzhou, China
Guangzhou, China

lllw1000@163.com

Abstract. This exploration aims to study how machine learning technology can identify student behavior and thus strengthen the supervision and management efficiency of educational venues. High-definition monitoring equipment was deployed in a middle school, and advanced sound capture technology was used with the data support of the learning management system to collect and study all the activities of students in detail. Using optical flow, short-time Fourier transform and random forest techniques, data was preprocessed, features were extracted and models were trained. The latest research found that the constructed model performed well in identifying collective activities, with an accuracy rate of up to 95% and an AUC value of 0.97; its performance was slightly insufficient in high-difficulty behavior recognition such as gesture recognition and complex behavior analysis. The research results show that although this model performs well in some specific behavior recognition tasks, it still needs further improvement and enhancement when facing more varied and complex behavior patterns. Future scientific research will focus on improving the model's wide adaptability and real-time processing efficiency to better meet the requirements of diverse educational environments.

Keywords: component; student behavior recognition; machine learning; feature extraction; behavior analysis

1. Introduction

In today's educational ecology, it is increasingly important to accurately grasp students' behaviors. This not only helps educators gain a deeper insight into students' learning progress, but also instantly optimizes the configuration of teaching plans and teaching resources. With the rapid advancement of machine learning technology, its application in behavior analysis and identification has highlighted unlimited possibilities and significant advantages. The purpose of this exploration is to master how to accurately identify students' behavior patterns with the help of machine learning technology, so as to optimize the teaching quality and environment. This study is committed to sorting out and processing behavioral data, screening and optimizing applicable machine learning models, in order to achieve accurate prediction and in-depth analysis of students' behavioral

M. R. Mohyuddin and N. A. D. IDE (eds.), *Proceeding of the 2024 International Conference on Diversified Education and Social Development (DESD 2024)*, Advances in Social Science, Education and Humanities Research 899,
https://doi.org/10.2991/978-2-38476-346-7_10

habits in various educational environments. Further research will focus on the actual performance of this technology in real teaching scenarios to confirm its expected effects and practical value. These exploratory works will build a scientific theoretical framework and a reliable practical foundation for the practical application of machine learning in identifying educational behaviors.

2. Student behavior recognition based on machine learning

2.1 Basic principles of behavior recognition

The key to behavior recognition is to predict or confirm the behavior pattern through detailed monitoring and in-depth analysis of individual behavior. The flowchart shown in Figure 1 shows this process in detail: First, the behavior is first reviewed, and then the next steps are planned based on the witnessed behavior characteristics. In the actual application of machine learning, it starts with collecting various types of information on student behavior, covering various forms such as video surveillance, sound recording or electronic monitoring of learning process. Then, the system will perform preliminary processing on the collected initial information, filter out interference signals, and extract key attributes. This link is crucial. After all, the quality of these attributes will directly determine the effectiveness of subsequent modeling work. Then, using the processed data, the machine learning model is used to identify various behavior patterns, such as student participation in class, lack of concentration, etc. In the process of model training and upgrading, the calculation method will undergo multiple cycles of improvement to achieve the highest recognition efficiency. After the model is trained, it can analyze emerging behavioral data in real time and predict students' potential behavioral trends. Once a behavioral pattern is identified, it may indicate learning difficulties or other issues that deserve attention. At this time, the corresponding countermeasures can be immediately initiated. This system forms a complete cycle, starting from collecting data, to identifying behavior, and finally to taking action. Its purpose is to promote students' learning progress and personal growth through accurate analysis of behavior[4].

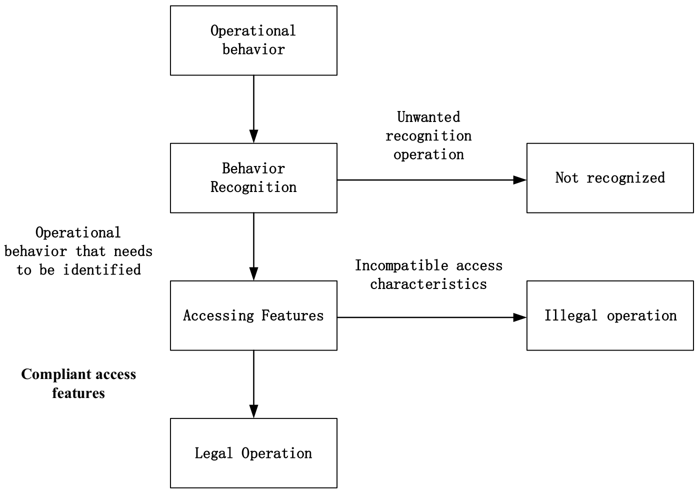

Figure 1. Identification process

2.2 Data Collection and Processing

The first step is to capture accurate information from various channels, and the second step is to thoroughly organize and optimize this information. First, in the classroom, high-definition monitoring equipment captures 30 images per second, while directional microphones capture audio at a frequency of 44.1 kHz to ensure the clarity of conversation information. For example, student activity information from the learning management system (LMS), such as login records and homework submission times, are marked with timestamps to mark specific user interaction moments. The OpenCV library is used to capture video data frames, and then the Gaussian mixture model is used to remove the background to extract the student's behavior dynamics. First, the digital audio data is quickly Fourier transformed to obtain its spectral characteristics. Then, high-pass and low-pass filters are used to remove noise at abnormal frequencies. Finally, the Mel frequency cepstral coefficient technology is used to extract the core attributes of the audio data. When processing digital interaction data, the timestamp is analyzed and the activity frequency is calculated. Then, the Z-score normalization method is used to eliminate the influence of different dimensions and improve the consistency of data comparison.

2.3 Selection and training of machine learning models

In the project of identifying student behavior, we carefully selected appropriate machine learning models and trained them in detail, which specifically involved the following unique and detailed operation process: For video data processing, especially student behavior patterns, we used a pre-trained convolutional neural network ResNet-50 model and adjusted it accordingly. This algorithm model effectively overcomes the performance degradation problem during deep neural network training with the help of residual learning architecture, can efficiently process complex image information, and accurately capture subtle differences in movements. In the field of audio data processing, bidirectional long short-term memory network (Bi-LSTM) is widely used. It is good at capturing contextual relationships, thereby optimizing the understanding and classification of students' voice behaviors. The network captures time-varying characteristics from audio features, such as students' enthusiasm for participating in classroom discussions. When honing these artificial intelligence models, the first step is to perform preliminary processing on the corresponding data set and divide it into eight parts: 20 split the data proportionally to distinguish two sets for training and testing. Adjusting the model parameters on the training data set with the help of the cross-validation method can effectively prevent the model from overfitting the training data. The cross entropy is used as a metric to determine the classification ability of the model. The Adam optimizer is used to adjust the weights. This method is not sensitive to the initial learning rate and is very suitable for dealing with large-scale data sets. In the entire model training process, after several rounds of iterative learning, the accuracy and error of the model will be tested on the validation set after each learning, and then the parameter settings will be adjusted and refined. This process continues until the performance of the model reaches a stable high level.

3. Application process

3.1 Application overview

This study set up a middle school with 1,200 students and 70 teachers in the suburbs as an experimental base. It is well-equipped and has rich backgrounds of teachers and students. It has created a unique experimental place for research in the field of machine learning. In the school, in order to accurately grasp the behavior patterns of students, 80 sets of high-definition monitoring equipment and 40 sets of highly sensitive recording equipment have been fully deployed. These high-tech instruments are cleverly placed in key locations such as classrooms, libraries, canteens and sports fields to ensure that students' behavior can be tracked in real time regardless of the occasion. Each student connects to the campus wireless network through a private handheld device and uses the learning management system (LMS) to record their learning process in detail, including the number of daily logins, online learning time, interaction frequency, and homework upload status. The amount of data generated each month is approximately 800GB, which includes detailed learning behavior records and time series information of student interactions. The researchers worked closely with the information technology department of the university to ensure that the data is fully complied with the current education data protection laws and policies from collection to transmission to storage. In addition, all students and parents participating in this study fully understood and agreed that their relevant information could be used for research projects in advance. In the process of data processing and analysis, we adhere to strict anonymization and encryption principles to ensure the security of personal information and prevent any form of leakage. This study aims to use machine learning technology to accurately capture and predict students' behavior patterns, provide educators with real-time, data-based insights, and help them improve teaching strategies and learning places.

3.2 Implementation steps

In the student behavior recognition system implemented in this middle school, the expert team used a variety of precise methods and technical processes to ensure the successful application of machine learning technology. The first phase of the engineering task is to collect massive data, covering 80 surveillance cameras placed in the classroom, which capture 30 pictures per second; at the same time, 40 microphones simultaneously capture the surrounding sounds. These machines and equipment can automatically transmit the collected data to the school's centrally managed data center, and then archive it on a daily basis. In the initial data processing stage, the video data was first screened to eliminate non-effective video records at night. Then, through the optical flow method, the formula $\mathit{Optical\ Flow}\left(I_x, I_y, I_t\right)$ The pixel changes between consecutive frames are calculated to capture the slight changes in the student's movements. The audio data is denoised by spectral thresholding, and then the short-time Fourier transform (STFT) is used, using the formula $X(k,\tau)=\sum_{n=0}^{N-1} x(n)w(n-\tau)e^{-j\frac{2\pi kn}{N}}$ Convert the signal from the time domain to the frequency domain for analysis. For interactive data extracted from the learning management system (LMS), a logarithmic transformation is used $y=\log(1+x)$ to reduce the skewed distribution and make the data more suitable for subsequent numerical calculations and research, in the feature extraction stage, we extract important features from the pre-processed audio and video data. With the help of convolutional neural network (CNN) technology, video information can accurately identify facial expressions and gestures. At the same time, the audio part focuses on exploring the rhythm and intensity characteristics of the sound, and uses Mel frequency cepstral coefficients (MFCC) as the main sound feature extraction tool. The researchers mixed various complex information with personal behavior data and used the random forest method to teach the computer how to distinguish. This training process divides the data into 80% and 20%: 20, and divides the data into training data sets and testing data sets. In the process of cultivating artificial intelligence models, cross-validation techniques are used to fine-tune the number and level of decision trees.

4. Effect evaluation

4.1 Evaluation method

This study aims to comprehensively examine the effectiveness of machine learning models in identifying student behavior in Maple Leaf School. A complete set of evaluation strategies is adopted, covering quantitative evaluation and model performance index measurement. Statistical indicators such as precision, recall rate, F1 score and ROC curve constitute important criteria for measuring the performance of the model in practical application. Accuracy is the percentage of correct and incorrect judgments of the artificial intelligence model, and the calculation method is as follows:

$$Accuracy = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP} + \mathrm{TN} + \mathrm{FP} + \mathrm{FN}}$$

Among these four abbreviations, "TP" can be understood as "True Positive Samples", which means correctly identified positive samples; "TN" refers to "True Negative Samples", which means correctly identified negative samples; and "FP" is "False Positive Samples", which means incorrectly identified positive samples; "FN" is "False Negative Samples", which means incorrectly identified negative samples. The recall rate reveals the accuracy of the model in identifying correct cases, which is crucial for not neglecting important behavior recognition in educational environments. It is calculated as follows：

$$Recall = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}}$$

The F1 score is the harmonic mean of precision and recall, which is used to measure the balance between model accuracy and retrospective ability. The calculation formula is:

$$F1 = 2 \times \frac{\mathrm{Precision} \times \mathrm{Recall}}{\mathrm{Precision} + \mathrm{Recall}}$$

On the basis of measuring the accuracy of the model, in order to deeply evaluate the classification effect of the model at different thresholds, we used the receiver operating characteristic curve (ROC) and the area under it (AUC value) as evaluation indicators. The ROC curve shows the relationship between the proportion of positive examples correctly identified by the model (true positive rate) and the proportion of positive examples incorrectly identified (false positive rate) under different classification judgment criteria, which intuitively reflects the performance of the model. The AUC value is close to 1, which means that the stronger the recognition ability of the model is, the more accurately it can distinguish between positive and negative samples.

4.2 Experimental results and analysis

The student behavior recognition experiment conducted in the middle school showed that the constructed model showed excellent discrimination performance. During the evaluation process, many performance indicators were comprehensively considered. After in-depth analysis, the practical application value of the model and its optimization areas were revealed in the educational environment. For various behavior recognition tasks, the model showed different accuracy, recall rate and F1 value. Especially in measuring students' enthusiasm for participation and team interaction, the model can complete the task more accurately, but in capturing subtle behavior patterns, such as students' body language (such as body movements), the performance is slightly inferior.

TABLE I. PERFORMANCE OF THE MODEL ON DIFFERENT CLASSIFICATION TASKS

Behavior Type	Accuracy	Recall	F1 score	AUC value
Student Engagement	92%	89%	90.5%	0.94
Gesture Recognition	85%	81%	83%	0.88
Collective activity recognition	95%	93%	94%	0.97
Voice Activities	88%	86%	87%	0.91

According to the data analysis in Table 1, the constructed model showed excellent accuracy and area under the curve (AUC) for the recognition of collective activities, highlighting its powerful effectiveness in capturing group behavior patterns. The display effect of gesture recognition technology is slightly inferior. This may be because the richness and complexity of gestures require the model to achieve higher accuracy. The experiment confirmed that the developed model has strong reliability and effectiveness in identifying student behaviors. Although there is still room for improvement in the identification of some specific behaviors, in general, the model can meet the basic requirements of educational scenarios for student behavior supervision and analysis.

4.3 Discussion

The general adaptability of the model is one of the focuses of discussion. The current model is mainly based on the information collected by a single school. This approach may restrict its adaptability in various teaching scenarios. Future academic exploration is expected to optimize the wide applicability and adaptability of the model with the help of data integration in a diversified educational environment. Current research has not yet deeply explored the potential of the model in real-time application. In practical application, the recognition system for student behavior must have efficient real-time data processing and analysis capabilities, which invisibly increases the challenges of algorithm improvement and computing resource allocation. Subsequent research should focus on improving the computing efficiency of the model and reducing the dependence on computing hardware to meet the needs of real-time supervision.

5. Conclusion

This experiment ran machine learning technology at Maple Leaf Middle School, proving the application prospects and effectiveness of student behavior recognition systems in educational scenarios. By carefully collecting and processing information, the model has achieved outstanding results in many behavior recognition tasks, especially in distinguishing group activities. There is still potential for improving the recognition accuracy of gestures and complex movements, which shows that our models and data processing methods need to be further improved. In the future, scholars can focus on enhancing the adaptability of the model in various educational environments, exploring faster data processing methods, and improving the model's skills in capturing subtle behavioral changes, and strive to create smarter and more flexible educational auxiliary tools, so as to more effectively support the teaching team and optimize students' learning places.

References

1. Li M. Recognition of Psychological Characteristics of Students' Behavior Based on Improved Machine Learning[J]. Journal of Sensors, 2021, 2021: 1-12.
2. Jingchao H, Zhang H. Recognition of classroom student state features based on deep learning algorithms and machine learning[J]. Journal of Intelligent & Fuzzy Systems, 2021, 40(2): 2361-2372.
3. Trabelsi Z, Alnajjar F, Parambil M M A, et al. Real-time attention monitoring system for classroom: A deep learning approach for student's behavior recognition[J]. Big Data and Cognitive Computing, 2023, 7(1): 48.
4. Abdallah T B, Elleuch I, Guermazi R. Student behavior recognition in classroom using deep transfer learning with VGG-16[J]. Procedia Computer Science, 2021, 192: 951-960.
5. Chonggao P. Simulation of student classroom behavior recognition based on cluster analysis and random forest algorithm[J]. Journal of Intelligent & Fuzzy Systems, 2021, 40(2): 2421-2431.