Construction of Learner Behavior Analysis Model in Immersive Virtual Reality

Based on Data Mining Technology

Hejin Wang[1(✉)] and Chengzheng Li[2]

[1] The University of Library, Sichuan University of Arts and Science, Dazhou, Sichuan, China
363366070@qq.com

[2] Modern Educational Technology Center, China West Normal University, Nanchong, Sichuan, China

Abstract. In the IVR environment, tracking and understanding learners' learning behavior is conducive to timely monitoring and guiding their learning status in the whole process, and is also conducive to system development and designers to further ensure the balance between teaching objectives and task settings. Based on data mining technology and the characteristics of IVR learning environment, this study discusses the deep integration of technology and learning data, constructs a learner behavior analysis model and applies it in the actual classroom of IVR. The specific distribution of learners in different clusters and the frequent sequence patterns of each cluster are found in practice, and the behavior sequence patterns of people with different performance levels can also be found by behavior sequence analysis. The results show that the behavior analysis method based on data mining technology can comprehensively reflect the learning state of students, provide the basis for teachers to implement accurate policies in IVR learning environment, and promote the data, scientization and precision of educational evaluation.

Keywords: Immersive · virtual reality · Learning behavior method · Data mining technology

1 Introduction

Immersive virtual reality (IVR) uses Head-mounted Display (HMD) to provide a virtual sense of reappearance of reality for participants' visual, auditory, tactile, olfactory and other senses [1]. Because it brings users an immersive virtual world, it is widely used in teaching scenes [2]. In addition, because of its special application environment, IVR has brought new breakthroughs to the related research in the field of social sciences, which can not only extract more accurate and important information, but also do not need to consider moral issues [3]. An analysis of the literature reveals a growing number of recent studies mining relevant learning data to explore the learning effects of participants in immersive learning environments [4, 5].

The application of data mining in the teaching cycle is known as educational data mining and is divided into three stages. In the first stage, the teaching environment is

X. Yuan et al. (Eds.): ICEKIM 2023, AHCS 13, pp. 960–971, 2023.
https://doi.org/10.2991/978-94-6463-172-2_101

adjusted according to the students' personal data before teaching, and the support for EDM is actively provided. In the second stage, the learning data are explained through EDM, and the relevant suggestions are put forward. In the third stage, the applied educational environment or form is evaluated [6]. At present, more and more researchers use EDM to analyze the behavior sequence data of learners in virtual environment, so as to explain the "why" and "how" of virtual environment to promote learners to learn better [7, 8]. For example, in the study, Cheng (2019) encoded the behavior of participants in chronological order by viewing the video of learning records, and proposed different behavior patterns of teacher-student interaction in the process of learning [2]. Behavior sequence analysis is a method of coding participants' behavior information to calculate whether there is significance in the conversion between codes [9]. It uses quantitative statistics to explain the transformation between behavior and action. The above research shows that the analysis of learning behavior based on data mining can evaluate learners' learning state in IVR environment more comprehensively and scientifically.

Based on the behavioral analysis method of data mining, this study constructs a model of learner behavior analysis in IVR and applies it to practice, aiming to track and understand the learning status of learners, which not only provides effective suggestions for monitoring and guiding the teaching process, but also provides effective guarantees for system developers and designers to further balance teaching objectives and task settings.

2 Analysis of Learner Behavior in IVR Environment

2.1 Analysis of Learning Behavior

Siemens points out in the study that learning analysis is a method of promoting understanding and optimizing learning and environment by collecting, measuring, analyzing and reporting data on learners and their environments [10]. Learning analysis can not only judge the learners who perform poorly in the learning process, but also carry out appropriate intervention according to the results of the analysis to provide personalized guidance for learners, so as to promote the realization of learning goals. The ultimate goal of learning analysis is to improve the efficiency of learning and teaching and to optimize the teaching effect. Therefore, learning analysis is of great significance to improve learners' learning effect [11].

The collection and identification of learners' data is a necessary prerequisite for the process of learning analysis. Fait divides the learning data into four dimensions: learning behavior data, learning network data, learning emotional data, and learning level data. The learning behavior data refers to the data between the learner and the learning environment or learning resources, including the operation action, the number of logins and so on. Then the specific behavior analysis tools are used to understand the learning state, path and behavior characteristics of learners.

2.2 Application of Learning Behavior Analysis in IVR

As data mining and processing technologies continue to evolve, researchers are increasingly interested in learning behavior analysis, which allows previously "invisible, neglected, and unprocessable" information to be effectively exploited and utilized. Visualization of valid data from the learning process not only reveals the potential clustering characteristics of different clusters of learners and explores different patterns of behavioral sequences, but also provides a comprehensive and systematic assessment of the characteristics and limitations of the IVR classroom environment, providing a reliable basis for educators' guidance and interventions, as well as helping researchers to reexamine and improve the functionality of the platform by recognizing the correlation between the learning environment and learning behavior patterns. It also helps researchers to re-examine and improve the platform features by recognizing the association between the learning environment and the learning behavior patterns, so as to find a balance between the pedagogical objectives and the game task setting aspects.

Therefore, a number of studies have confirmed the validity and reliability of learning behavior analysis in IVR, while the reliability has been verified through experimental data. This shows that the application of learning behavior analysis method in IVR education and teaching has certain research value.

3 Construction of Learner Behavior Analysis Model in IVR

3.1 Model Construction

In this study, the learning behavior of learners in IVR environment is taken as the research object, and the analysis model of learners' behavior in IVR is constructed, as shown in Fig. 1. The application of this model provides scaffolding for classroom education and teaching research in immersive virtual reality environment, improves students' learning effect in IVR learning environment, and improves classroom teaching efficiency. The model of this study mainly includes nine stages: problem definition, teaching equipment configuration, data collection, data preprocessing, behavior data mining, result visualization, result analysis, optimization of learning process, innovative teaching application and resources.

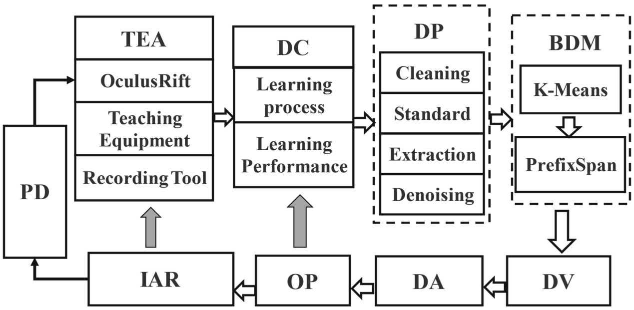

Fig. 1. Learning behavior analysis model

3.2 Model Connotation

PD: Problem Definition. At this stage, we should make clear the specific teaching problems that need to be solved in teaching research, and set up the direction mark for the next step of design.

TEA: Teaching Equipment Allocation. According to the problems to be solved, select the teaching resources, teaching equipment and teaching mode that accord with the application scene and the research object, and configure the software or tools needed for data collection.

DC: Data Collection. Including Learning Process and Learning Performance. In the data collection phase, the teaching characteristics and data that need to be used in the research process are collected or recorded through specific tools to meet the next data processing.

DP: Data Preprocessing. In the original data, only a small amount of information may be related to the target problem, and the ultimate purpose of data preprocessing is to obtain an effective and standardized data training set. This stage includes data cleaning, standardization, transformation, feature extraction and denoising.

BDM: Behavior Data Mining. Behavior data mining is to select appropriate behavior analysis tools or methods to deal with data information, so as to obtain interpretable and meaningful behavior analysis results.

DV: Data Visualization. In the visualization result stage of behavior analysis, implicit and effective interpretable educational information is found from the data results.

DA: Data Analysis. In the stage of data analysis, the differences between the data are explained from the point of view of education, and the influence of different data on learning results is explained.

OP: Optimization Process. The study provides a reliable basis for the guidance and intervention of educators in the teaching process through the similarities and differences of learners' behavior performance with different characteristics in IVR fire safety education class, and also plays an important reference role for learners to adjust and monitor the learning process.

IAR: Innovative Applications and Resources. In this stage, different behavior sequences and behavior ratios are used to help system development and designers re-examine and improve the functions of the platform in order to find a balance between teaching objectives and game task settings.

4 Practical Application of Learner Behavior Analysis Model in IVR

4.1 Experimental Design

The experiment was conducted in an information technology classroom in a middle school. Participants studied in the Fire Safety Education Game "Fire Safety Laboratory", a fire safety education game developed by Oculus Rift and IMP Studios. They were required to complete a series of operations in the designated exploration scene. The exploration time is 10 min per person.

4.2 Experimental Process

Problem Definition

In this experiment, the whole learning process of learners in exploration is recorded in the form of video, and then the potential behavior characteristics are obtained by using data mining technology, and the similarities and differences of learners' behavior of different characteristics in IVR primary and secondary school fire safety education class are analyzed.

Teaching Equipment Allocation

The configuration of teaching equipment in this experiment is divided into three parts, including the realization equipment of immersive virtual reality environment, the running equipment of teaching resources and the tool of recording learners' behavior data. The OculusRift developed by Oculus Company is selected to realize the realization equipment of immersive virtual reality environment, which is composed of HMD, space sensor and interactive control handle, which can track the position in a limited area. The control handle realizes the interaction with environment and objects in virtual reality three-dimensional space by moving in the real world through the different functions of sensors and control panels.

Teaching resources run the device, and this experiment consulted the PC configuration requirements for the retail version of the OculusRift virtual device published by Oculus in order to ensure the fluency of the virtual device as well as the teaching environment. In order to record a large amount of behavioral data from the learners during the exploration process, combined with the special learning environment, this study decided to use the EV recording tool to record the learning process. Video analysis, although leading to more workload, enables accurate coding of the learners' learning behavior, which helps the study to make more in-depth qualitative judgments and analysis.

Data Collection

In order to understand the behavior characteristics and laws of different learners in IVR fire safety education class, this experiment collected a large number of information feature data from the learning process, including learning behavior characteristics and learning performance characteristics. The characteristics of learning behavior refer to the behavior information data of learners. The behavior performance of learners in the whole learning process is recorded by recording tools. The information such as the type of each behavior and the time of behavior occurrence is recorded in the form of video, and then all the behaviors of learners in the video are encoded and processed in turn according to the behavior coding framework combined with a fixed time span. Learning performance characteristics refers to the learning through IVR fire education class, each student in the examination module to obtain the test results data, at the same time, when collecting the data, we must ensure that each learner's performance data and their behavior characteristics maintain an one-to-one corresponding relationship. In this experiment, a total of 59 learners recorded videos.

Data Preprocessing

All videos are imported into the Nvivo video analysis tool for video transcription, the purpose of which is to transcribe the information in the video into machine-recognizable data. The length of each video is 10 min, and then every 2.5 s is set as the time span according to the length of a single action of the learner in the video, and each small segment of the event is classified into a predetermined behavior label group in sequence. The behavior label is shown in Table 1, and a total of 14160 behavior codes are obtained. Considering that the long time repetitive action affects the effectiveness of the whole sequence pattern, this experiment combines the two action codes into a single action block to reduce the problem of more granularity of data caused by a large number of the same type of action, and this step is more conducive to the further interpretation of the sequence.

Behavior Data Mining

First of all, the two-stage clustering analysis method is used to classify the possible

Table 1. Learning Behavior Coding Framework

Coding	Behavior	Description	Example
MO	monitoring of activities on market level	Ask or view a time, task, etc.	Learning to explore the behavior of viewing tasks
EX	Exploration	Participants' touch and search for items in the environment	Looking for fire extinguishers
TH	Reflection	Pause, observe the environment	The act of stopping in the environment.
SH	Ask for help	Ask the teacher for help when you are in trouble	Raise your hand and ask for help
TR	Extinguish fire	Try to put out a fire	Pick up the fire extinguisher and open the ring
IM	Execution	Perform a series of fire extinguishers preparation tasks	Pull down the switch
AD	Compensation	To adjust the position, state of a person or object.	The act of making corresponding changes after an invalid operation.

different learning groups, as shown in Fig. 2. In the first stage, the appropriate clustering number is determined by the outline coefficient; in the second stage, the K-means clustering analysis is carried out according to the determined clustering number. This method focuses on the in-depth analysis of learners' potential behavior characteristics, which can help teachers, researchers and system designers to evaluate learners' behavior, and help to examine the characteristics and limitations of the educational platform. After analyzing the behavior characteristics of each cluster, the behavior codes of each student in all clusters are arranged in chronological order. According to the behavior code data, the behavior sequence is further analyzed. By calculating the frequency of each behavior code converted to another behavior code, the behavior frequency conversion matrix, conditional probability matrix and expectation matrix are obtained, and then the adjusted residual table is inferred to form the behavior transformation diagram. In order to more accurately analyze the sequential relationship between the behaviors.

According to the final score, the learners were divided into high score group and low score group, the top 27% of the students were divided into high score group, and the latter 27% of the students were divided into low score group, as shown in Fig. 3. Through statistics, it is found that the high score is the top 16 learners, a total of 3540 behavior codes, and the low score is the last 16 learners, a total of 3216 behavior codes. Then the PrefixSpan algorithm is used to mine the behavior sequence pattern of high packet data set and low packet data set.

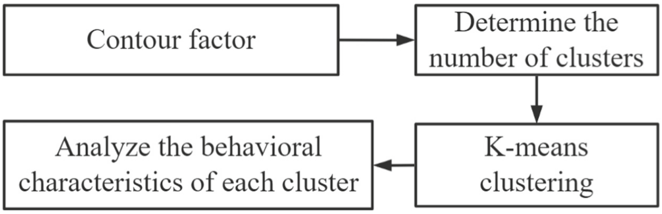

Fig. 2. Cluster Analysis

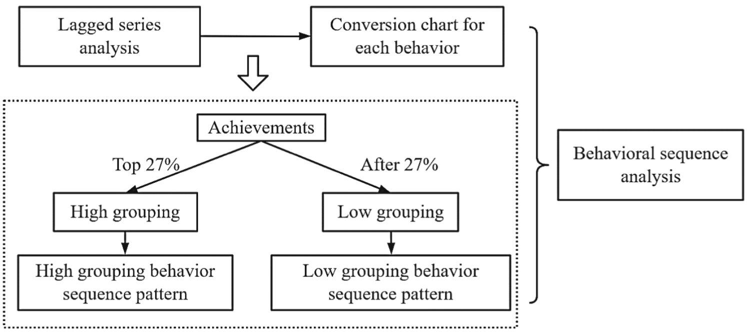

Fig. 3. Learning behavior analysis model

Data Visualization

Visualization of Cluster Analysis Data. Calling the K-Means clustering program algorithm of Python, the 59 groups of behavior sequence encoded data sets are packaged and imported, and the optimal clustering number is $K \leq 3$ according to the results of contour coefficients. After the specific samples contained in different clusters are obtained, the single factor variance analysis of multiple groups of samples is carried out by SPSS software, and the frequency of different behaviors in each sample is calculated respectively. The survey found that in cluster 1, the number of learners who passed the cluster accounted for 37.04%; in the second middle school, 75% of the learners passed the cluster; and in cluster 3, the number of learners who passed the cluster accounted for 91.67%. The behavior transformation is then described according to the behavior residual table of the three clusters, as shown in Fig. 4.

Visualization of Behavior Sequence Analysis Data The PrefixSpan algorithm is used to mine the behavior sequence patterns of high packet data sets and low packet data sets, and the support range is set to 80%. Frequent sequence patterns for high and low packets are shown in Table 2 and Table 3.

Analysis of the Data

Cluster Analysis From the results of cluster analysis and lag sequence analysis, it is found that there are two types of learners showing obvious learning characteristics. First, some learners are in a more cautious state in a strange learning environment, lack of in-depth thinking and effective cyclic action sequence transformation. From the perspective of behavior frequency, most of them are in a state of repeated thinking or stagnation, and lack of important problem-solving strategies to seek help. Immerse yourself in the cycle of shallow cognition Second, some learners are familiar with the environment through exploration in the learning process, and are used to solving problems quickly in the form

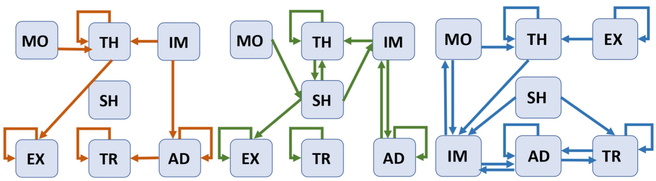

Fig. 4. Behavior Conversion Chart

Table 2. High packet behavior sequence pattern

Sequence ID	Sequence pattern	Support degree
1	TH-TH	90.78%
2	AD-IM-AD	86.24%
3	TH-TR-AD	83. 37%
4	MO-TH	81. 23%

Table 3. Low packet behavior sequence pattern

Sequence ID	Sequence pattern	Support degree
1	TH-TH	100%
2	TH-TH-EX	94.64%
3	AD-AD	89.17%
4	TH-EX-EX	82.09%
5	TH-AD	80.16%

of asking teachers for help when they encounter difficulties. This kind of learners show impatience or lack of in-depth thinking process.

Cehavior Sequence Analysis From the analysis of the results of the high group behavior sequence pattern, the high group learners are good at using the monitoring behavior such as viewing task list and asking task time to adjust and evaluate the task status, and further clarify their learning tasks and goals in order to improve the learning efficiency. In addition, they tend to use circular solutions to promote the achievement of goals and tasks by constantly adjusting their own directions or methods in a timely manner. This self-timely feedback method can help them quickly locate problems and identify all kinds of error operations.

Discussion

Learning Behavior Analysis Can Comprehensively Reflect Learners' Learning State in IVR Learning behavior analysis not only identifies potential clustering characteristics of learners in different clusters and explores different patterns of behavioral sequences, but also provides a comprehensive and systematic assessment of the status of learners. In this study, K-means clustering algorithm is introduced to divide learners into different clusters with their own characteristics, and then the adjusted residual table is inferred by lag analysis and calculation of behavior frequency transformation matrix, conditional probability matrix and expectation matrix, and then the behavior transformation diagram is formed to analyze the sequence relationship between different behaviors and the behavior differences among different clusters. Then, PrefixSpan algorithm is used to explore the behavior sequence patterns of high and low groups, and the behavior characteristics and laws of learners with different performance levels are analyzed and summarized, which provides visual and effective data for researchers to track and understand the learning situation of each student.

The Analysis of Learning behavior provides the basis for Teachers' effective Countermeasures in IVR Teaching Environment.

Teachers can analyze and implement strategies by monitoring the behavioral representations of learners in this environment. The results show that some learners are in a more cautious state in unfamiliar learning environments lack of in-depth thinking and the transformation of effective cyclic action sequences. From the perspective of behavior frequency, most of them are in a state of repeated thinking or stagnation. At the same time, they lack the important problem-solving strategies to seek help, immersed in the cycle of self-shallow cognition. For this kind of learners, teachers should actively guide and encourage learners to explore learning scenes and find solutions, understand the difficulties and conditions of students, in order to collect more clues to promote them to complete their learning tasks. Some learners are familiar with the environment through exploration in the learning process, and are used to solving problems quickly in the form of asking teachers for help when they encounter difficulties. This kind of learners show impatience or lack of in-depth thinking process. Teachers can encourage this kind of learners to stop and seriously analyze, reflect and adjust their problem-solving strategies when needed, and give appropriate guidance in order to achieve deeper reflection to improve learning efficiency.

Analysis of Learning behavior promotes the Scientific and effective Application of IVR in Education.

In this study, the potential behavior characteristics are obtained by using data mining technology, the similarities and differences of learners' behavior performance in IVR with different characteristics and the differences in behavior patterns of different learning performance are analyzed, and the characteristics and limitations of IVR course are comprehensively and systematically evaluated, which provides a reference for researchers to understand the relationship between situational learning environment and learning behavior pattern. It also helps IVR system development and designers re-examine and improve the functionality of the platform to find a balance between teaching goals and game task settings.

5 Conclusion

According to a relevant meta-analysis, the use of virtual reality technology in education has a positive impact compared to the traditional classroom [12]. It not only enhances learners' fun in terms of interactivity, immersion and engagement, but also reduces the risk of training learners' skills in hazardous situations under special teaching needs. In the IVR teaching environment, strategic assessment and effective intervention of learners' behavior is an important way to promote their efficient learning. This study clarifies the feasibility of learning behavior analysis in IVR scenarios through data mining techniques, and the analysis of learners' behavior has a significant effect on the improvement of teaching effectiveness.

Applying learning behavior analysis to IVR classroom teaching can not only determine learners who are underperforming in the learning process, but also implement appropriate interventions based on the analysis results to provide personalized guidance to learners, thus facilitating the achievement of learning goals. Of course, given the preliminary research results so far, a series of problems and challenges need to be faced in order to integrate data mining-based learning behavior analysis into the IVR curriculum system for popularization and normalization.

Future research can address the efficiency of behavioral data collection and analysis in education and teaching through intelligent technologies and means, strengthen the development of intelligent data collection technologies and analysis tools, such as artificial intelligence and 3D behavioral recognition technologies, integrate multidimensional behavioral data for learning analysis, and further explore the relationship between teaching effectiveness improvement and IVR classroom.

Acknowledgements. The work in this paper was financially supported by the project: Sichuan University of Arts and Sciences Research Start-up Fund. (Item No. 2022QD063).

References

1. Birenboim, A., Dijst, M., Ettema, D., de Kruijf, J., de Leeuw, G., & Dogterom, N. (2019). The utilization of immersive virtual environments for the investigation of environmental preferences. Landscape and Urban Planning, 189, pp. 129-138.
2. Cheng, K. H., Tsai, C. C. (2019). A case study of immersive virtual field trips in an elementary classroom: Students' learning experience and teacher-student interaction behaviors. Computers & Education, 140, pp. 103600.
3. Parsons, T. D., Gaggioli, A., & Riva, G. (2017). Virtual reality for research in social neuroscience. Brain sciences, 7(4), pp. 42.
4. Abdullah, M. E., Masek, A., Shamshuddin, M. K., Bahrudin, I. A., & Bujang, I. Z. (2016). A comparative study on the effect of virtual field trips (VFTs) through video aided learning (VAL) and traditional learning approaches on students knowledge acquisition. Advanced Science Letters, 22(12), pp. 4036-4039.
5. Kobayashi, K. D. (2017). Using flipped classroom and virtual field trips to engage students. HortTechnology, 27(4), pp. 458-460.
6. Peña-Ayala, A. (2014). Educational data mining: A survey and a data mining-based analysis of recent works. Expert systems with applications, 41(4), pp. 1432-1462.
7. Cheng, K. H., & Tsai, C. C. (2014). Children and parents' reading of an augmented reality picture book: Analyses of behavioral patterns and cognitive attainment. Computers & Education, 72, pp. 302-312.
8. Hou, H. T. (2015). Integrating cluster and sequential analysis to explore learners' flow and behavioral patterns in a simulation game with situated-learning context for science courses: A video-based process exploration. Computers in human behavior, 48, pp. 424-435.
9. Bakeman, R., & Quera, V. (2011). Sequential analysis and observational methods for the behavioral sciences. Cambridge University Press.
10. Siemens, G., & Baker, R. S. D. (2012). Learning analytics and educational data mining: towards communication and collaboration. In Proceedings of the 2nd international conference on learning analytics and knowledge, pp. 252–254.
11. Wu, F., Huang, L., Zou, R. (2015). The design of intervention model and strategy based on the behavior data of learners: A learning analytics perspective. In International Conference on Hybrid Learning and Continuing Education. Springer, Cham. pp. 294-301.
12. Chauhan, S. (2017). A meta-analysis of the impact of technology on learning effectiveness of elementary students. Computers & Education, 105, pp. 14-30.