ingegneria sismica
International Journal of Earthquake Engineering
Anno XLIII - Num. 1 - 2026

Learning Analytics-based Learner Behavioral Pattern Mining and Intervention Strategies for Online Programming Course Learners

Yang He[1,*], *Dun Mao*[1] *and Jiayu Li*[1]

[1] School of Mathematics and Physics, Fuzhou Preschool Education College, Fuzhou, Jiangxi, 344000, China

SUMMARY: *The learning time and location distribution of students in online learning has dynamic characteristics, this paper proposes the concepts of time entropy and location entropy to quantitatively analyze the temporal and spatial distribution of students and their changing characteristics. The regularity analysis of online learning behavior is carried out by using actual entropy, and the graph model is used to model the learning process data. Through online learning behavior clustering, the learning status of different categories of learners is mined. Based on real online learning data, the behavioral patterns of online programming course learners are explored, and corresponding learning behavior intervention strategies are proposed based on the research results. The results show that "anytime" learning usually represents a higher level of learning engagement and a greater chance of achieving good learning outcomes. However, "anywhere" learning is negatively correlated with learning outcomes, i.e., frequent changes in learning location are not conducive to good learning outcomes. Based on the online learning behaviors, the students were clustered into high-involvement, random-engagement and low-involvement types, and their mean scores in the regular experiments were 90.94, 80.02 and 70.23 respectively, and their mean scores in the final exams reached 88.03, 76.07 and 64.12 respectively, which indicated that those who performed well in the final programming test had already shown some advantages in the regular experiments.*

KEYWORDS: *online programming learning; spatio-temporal distribution law; actual entropy; graph model; cluster analysis; learning behavior analysis*

1 Introduction

In recent years, with the advancement of education digitization and the popularization of the Internet, the public's demand for online education is also growing rapidly, and it is estimated that the online education industry will maintain steady growth, the market size will reach 590.19 billion yuan, and the size of the user base will reach 352 million people [1]. With the rapid development of online education, its problems have become more and more prominent, for example, the online teaching method makes it impossible for teachers to monitor the learners, so it is difficult to give learners personalized guidance [2]. The increase of online learning resources, educational software, and the extensive use of the Internet in education has created a large amount of educational data, which records the trajectory of the learner's learning behavior, and how to make good use of this educational data in order to gain a deeper understanding of the learner's learning situation is a major challenge facing the current [3-5].

*hydwyyx2025@163.com
https://doi.org/10.65102/is2026047
Published: 30 April, 2026

However, the existing research mainly focuses on the use of "black-box" methods to draw direct conclusions, which is difficult to provide information about the learning process.

In addition, the arrival of the new generation of artificial intelligence era has a profound impact on people's learning and life. Cultivating innovative talents that meet the needs of intelligent society requires good computational thinking, programming ability and deep knowledge of intelligent society [6]. At home and abroad, governments have introduced a series of educational policies and taken positive measures for programming education to respond to the call for talents in the intelligent era [7]. Programming education has become an important part of the education system at home and abroad. For example, the United States in the early 2000s to the present, programming education and K12 education has been fully integrated, social multi-party forces to cooperate to promote the scientific systematization and diversification of programming education, Codecademy, Khan Academy, edX, Coursera and other online education platforms provide a large number of programming courses and learning resources [8]. The United Kingdom in its 2014 version of the National Curriculum Standard included computer courses into the compulsory courses for students aged 5-16, in which programming education occupies an important position [9]. In China 2019, it is clearly proposed to promote the promotion and popularization of artificial intelligence and programming education in primary and secondary school stages. It can be seen that programming education has become an important way to cultivate talents, and under the continuous promotion of education reform, how to improve the effect of students' programming learning is an urgent problem to be solved at present.

It has been shown that learning analytics helps the researcher to understand the effectiveness of learning by measuring and analyzing the data of students' learning process so as to optimize the learning process [10]. Learning analytics refers to the measurement, collection, analysis, and aggregation of data about students and their contexts in order to understand and optimize the learning process and learning environment [11]. Research on instructional intervention strategies based on students' behavioral patterns is an important research direction in learning analytics research. Previous studies have shown that analyzing the sequence of students' learning behaviors can help instructional stakeholders to grasp the potential behavioral patterns of students, and to discover the differences in learning behaviors among different groups of learners, throughout the stages of learning activities, and at different stages of learning [12-15]. By identifying and monitoring the key behavioral sequences that affect students' learning outcomes, it is beneficial for teachers to provide students with more targeted and timely implementation of instructional interventions based on data analysis in order to improve learning outcomes.

In the study of learner behavioral pattern mining, literature [16] used betty's brain virtual agent to teach students, capturing a large amount of data about student learning interactions in the process, and using data mining methods to assess and compare student learning behaviors from these interaction trajectories. Literature [17] used k-mean clustering of informational data on 700 students using Aleks, a Web-based, adaptive assessment and learning system, and, using this approach, was able to identify five different profiles from these students, and then monitor students in need of help based on these profiles for Aleks, and design and implement appropriate interventions for groups of students with similar characteristics Interventions. Literature [18] used the DSM algorithm to study how students collaborate and the learning behaviors they exhibit when learning through modeling tasks during group learning, as well as teaching virtual agents in the betty's brain system, where the DSM algorithm uses a combination of the SPAMc algorithm and plot mining to mine students' learning behaviors. Literature [19] proposes an "adaptive learning" online teaching approach, which adapts to the learning needs of students as they begin to progressively learn the course content, thus providing a customized experience based on prior knowledge and effectively helping learners to master the content. Literature [20] proposes a collaborative learning (CL) and initiation learning as well as self-regulated learning (SRL) and feedback online teaching approach based on developing students' collaborative skills and regular study habits to improve students' computational skills in website design.

Programming course learning behavioral pattern mining research aims to mine students' learning patterns from their programming behavioral data, which can help educators to better understand students' programming learning process, and also help students to adjust their own patterns towards better learning patterns [21]. This type of research first collects and preprocesses students' programming behavior data, extracts behavioral features from them, and then mines and analyzes students' learning patterns using different methods on this basis. According to the different methods of obtaining learning patterns, it can be divided into two categories of research: educational data mining methods and other methods. In the study of educational data mining methods, literature [22] used students' programming behavioral features as observations for cluster analysis, and classified students' behaviors in programming learning into effective, resilient, and ineffective behaviors by analyzing the statistical differences between the clustering results and the learning outcomes. Literature [23] calculated the probability mean of two students' code snapshot sequences transforming into each other as a distance by using Hidden Markov Model (HMM), and based on this distance 3 clusters were obtained by clustering the students' programming patterns using K-Means, and there were differences in students' performance in different clusters. Literature [24] explored the differences in programming behavior patterns between groups of students with different achievements based on student compilation data, the study was conducted based on compilation run data during student programming, the main object of the study was the transition paths between 11 different programming states, the different paths represent the students' programming learning patterns, and the students were divided into 3 classes with different patterns by using statistical methods. Literature [25] proposes to study students' programming learning mode by analyzing the changes of learning curves in the process of students' programming, which considers the node types of the abstract syntax tree (AST) of the code as the programming knowledge points, and constructs four knowledge point models through the actual programming data of the students, and generates the learning curves from the models to represent the different learning modes of the students. Literature [26] proposes a hybrid machine learning approach, which combines K-Means clustering with enhanced neural network (BNN), enhanced by AdaBoost engine enhanced Extreme Learning Machine (ELM), aiming to mine and analyze the behavioral patterns in the Java programming course can reveal key information about learning styles, performance gaps and content interactions.

Some studies have extracted features with the help of other methods or tools in order to achieve mining analysis of learning behavior patterns. For example, literature [27] collected features such as total programming time, number of unique assignments attempted, number of correct submissions, number of incorrect submissions, number of submissions that could not be compiled, total scores of all submissions, and total number of changes made in the code to characterize the learning behaviors of students in programming coursework. Literature [28] collected and analyzed the log data of students' learning activities using a learning management system and an automated assessment tool in order to understand the approach to learning programming, and it was found that students have certain tendencies when learning programming, such as a tendency to use slides that can support copying, pasting, and searching, a tendency to watch fewer videos while completing the programming tasks, and a tendency to go to the course forum for help only when they encounter problems . Literature [29] used clickstream data to collect and analyze students' behaviors in Scratch programming, and used clickstream analysis to understand students' engagement and learning patterns in a graphical programming environment, and the results of the study showed a strong correlation between students' behaviors and their performance in programming tasks. Literature [30] mapped the three-dimensional framework of computational thinking and Bloom's newly revised cognitive categorical goals as a framework for evaluating computational thinking, collected and analyzed students' screen activities and communicative voice data during programming, and the results of the study showed that students' cognitive levels were most at the stage of understanding and analyzing problem-solving steps and code modification, followed by the stage of applying programming concepts to creating code, and finally the evaluating code with optimization stage. Literature [31] collected data on learners' programming behaviors as they completed programming tasks in the Online Python programming environment, which consisted of snapshots of learners' code as they ran (piloted) their programs, as well as patterns of their behaviors during programming such as the number of tasks selected, the number of overall code runs, the average number of code runs per task, the number of successful tasks, and the amount of time spent on each task. Literature [32] used an anomaly detection method to predict dropouts by clustering the data generated in a programming course using unsupervised learning based on information obtained from educational data mining.

This paper proposes an entropy-based quantitative analysis method of spatio-temporal distribution to quantitatively represent the characteristics of students' spatio-temporal distribution, and classify students based on temporal entropy and spatial entropy. By analyzing learners' online learning behaviors, the corresponding actual entropy values are defined and calculated to assess the regularity of individual learning behaviors. A method of serializing learning behavior data and a similar behavior clustering algorithm based on graph model are designed to realize learning state clustering analysis. In the datasets of two different learning contexts, different spatio-temporal behavior patterns of students are identified, and the characteristics of different spatio-temporal behavior patterns and the relationship with learning effectiveness are analyzed. Through the behavioral dynamics theory, a behavioral dynamics model of learner-resource interaction is established. Combined with the simulation results of the model, the influence of teachers' online announcement behavior on students' resource learning behavior is verified. The student portraits are clustered into different categories, and the degree of students' mastery of each type of knowledge points is analyzed from the group portraits. Synthesize the results of the study and propose learner behavior intervention strategies in online programming courses.

2 Learner behavior pattern mining method based on online learning behavior analysis

2.1 Entropy-based quantitative analysis method of spatio-temporal distribution

In online learning, students can access course materials and interact with peers or the instructor “anytime, anywhere”. Students can learn through online peer support mechanisms and often have more opportunities to find answers to their questions than in a traditional classroom. In addition, students will take more responsibility for self-regulated learning. Location-based mobile learning (LBML) allows students to interact with both physical and virtual environments. Specifically, “anytime, anywhere” is seen as a major advantage of online education, but it remains unknown how students can utilize temporal and spatial convenience for learning purposes. In addition, the frequent switching of learning environments and the fragmented nature of learning time and content have raised doubts among some educators about the quality and effectiveness of fragmented online learning.

Online learning has gradually got rid of the limitations of accessing through wired networks and fixed terminals, truly realizing the state of learning anytime and anywhere. However, while learning has become more and more convenient, it has also produced some randomness and uncertainty in the timing and location of learning activities, large fluctuations in the duration of learning, and frequent changes in the learning location and other characteristics. In online learning, the distribution of students' learning location has dynamic characteristics, students may visit online learning platforms at different locations with arbitrary frequency; students' learning time also has randomness, students may carry out learning at any moment of any day. “Entropy” is usually used in informatics to indicate the randomness and uncertainty of events. The spatial and temporal distribution characteristics of students coincide with the characteristics of entropy, and if students study in a certain location or time period as a random event, the entropy can be used to measure the distribution characteristics of learning behavior in multiple learning locations and learning time periods. When the value of entropy is larger, students' learning locations or learning periods are more scattered; when the value of entropy is smaller, students' learning locations or learning periods are more concentrated. In order to quantitatively analyze how students learn anywhere and anytime, this paper uses the concept of entropy to characterize the spatial and temporal distribution of students.

The concept of entropy was originally derived from thermodynamics to describe the behavior of the elements of a thermodynamic system, and it is related to the amount of order or disorder in the system. With respect to machine learning, entropy is a measure of randomness in the information being processed. For example, decision tree algorithms use entropy to determine the optimal branching threshold for nodes in modeling. In educational research, entropy has been widely used in learning analytics and educational data mining research.

The formula for entropy is as follows:

$$Entropy = -\sum_{k=1}^{n} p_k \log_2 p_k \tag{1}$$

where p_k is the discrete probability distribution of the discrete random variable X, indicating the probability of the event occurring.

In this paper, the location entropy and time entropy values are used to measure the concept of “anytime anywhere” learning, i.e., the magnitude of the entropy indicates the degree of concentration or dispersion of the learning location and time distribution.

2.1.1 Positional entropy calculation

In learning activities, there is uncertainty in the distribution of students' locations. The probability that a student learns at a certain location depends on the number of times he/she learns there and the number of times he/she learns at all locations. Therefore, according to the formula for entropy (1), the random event X is defined as a student carrying out a learning activity at a certain location, and the probability that X occurs is determined by the ratio of the number of times that the student learns at that location to the total number of times that the student learns. The uncertainty of a student's learning location is related to the number of locations where the learning activity occurs.

In this paper, we use “location entropy” to represent the concept of “learning anywhere”. In this paper, we categorize all the locations of a student to get the number of locations and the number of times each student learns at each location. The location entropy of student S is calculated by the formula:

$$LocationEntropy_S = -\sum_{i=1}^{n} p_i \log_2 p_i \tag{2}$$

where n is the number of location categories of student S and p_i denotes the probability of learning at location i, which is calculated as follows:

$$p_i = \frac{\text{Number of learning sessions at position } i}{\text{Total number of learning sessions for student } S} \tag{3}$$

Location entropy measures the distribution and variation of students' learning locations. The larger the value of location entropy, the more uneven the distribution of geographic locations when students use the client to study or there are more different categories of study locations; the smaller the value of location entropy, the fewer the categories of geographic locations when students use the client to study, and the more fixed the study locations. When the location entropy of a student is 0, it means that the student always uses the mobile to study in the same geographic location.

2.1.2 Time entropy calculation

In learning activities, there is uncertainty in the distribution of students' time. The probability that a student learns in a certain time slot depends on the number of times he/she learns in this time slot and the number of times he/she learns in all time slots. Therefore, according to the formula (1) for entropy, a random event Y is defined as a student carrying out a learning activity in a certain time slot, and the probability of Y occurring is determined by the proportion of the number of times that the student learns in that time slot to the total number of times that the student learns. The uncertainty of a student's learning time is related to the number of periods in which the student's learning activities occur.

In this paper, we use the concept of "time entropy" to represent the concept of "anytime learning". Temporal entropy measures how students allocate their learning time, and it indicates the degree of concentration and dispersion in the distribution of students' learning time. For example, when A has a greater time entropy than B, it means that A uses the LMS to study more often than B. The formula for the temporal entropy of student S is:

$$TimeEntropy_S = -\sum_{j=1}^{m} q_j \log_2 q_j \tag{4}$$

where j is the number of time period categories, and q_j denotes the probability of learning in time period j, which is calculated as follows:

$$q_j = \frac{\text{Number of study sessions within time period } j}{\text{Total number of study sessions for student } S} \tag{5}$$

According to this formula, if a student's temporal entropy is 0, the student always accesses the LMS during the same time period.

2.2 Regularity analysis of online learning behavior

In the current research environment, quantitative analysis of the relationship between learners' behavioral styles and learners' academic performance is an important step in personalized education. Experiments have demonstrated that some learners' personality traits (e.g., responsibility, sincerity, etc.) are strongly correlated with academic performance, especially conscientiousness. In the experiment it can be described as follows: the correlation of academic performance is the largest. Therefore, the regularity of online learning is used as an important dimension to predict students' academic performance in order to increase the accuracy of the model.

In order to measure the relationship between the regularity of students' online learning and academic performance, two methods are designed in this paper: designing the student login time difference scatter model and designing the actual entropy function.

(1) Scatter model of student login time difference

In order to portray the regularity of students' login learning time, this paper proposes a time difference scatter model between two logins in close proximity: the time interval between every two logins of a student is calculated separately, plus the number of logins of a student in an experimental cycle, the distribution of the student's online learning time can be portrayed; the learner's learning time is mapped to a two-dimensional coordinate system for visualization, and the user's learning time is made into a Scatterplot. As a result, the maximum, minimum, and average time intervals of each user's login interval and the truncated mean with the maximum and minimum values removed are counted, and their relationship with grades is analyzed. The one with the highest correlation is chosen to be added to the model to improve the prediction accuracy of the model.

(2) Design of the actual entropy function

Assuming that the learning time period is calculated on a weekly basis, and that assignments are submitted on a weekly basis, this paper divides the week as a time period by dividing the week into seven equal portions, each spanning one day, with the week coded from 1 to 7 (e.g., Monday is 1, Tuesday is 2, and so on). The learner's time for each login to learn is then mapped to this discrete time series $(1,2,3,4,5,6,7)$, or if the time distribution spans across two days, then this time is mapped to the longer time series. The actual entropy function is used to measure the coefficient of temporal regularity of user login learning, as defined in equation (6):

$$S_{\varepsilon} = x^2 \left(\frac{1}{n} \sum_{i=1}^{n} \Lambda_i \right)^{-1} \ln n \tag{6}$$

where x represents the number of weeks occupied by the mapped time series; n represents the number of mapped sequences, in order to accurately extract the learning regularity of the learner, here n must be taken as a consecutive sequence; and Λ_i represents the length of the shortest sequence that has not appeared before at the beginning of the i th.

If the length of time occupied by the sequence is shorter, the effect of entropy is more obvious. If the actual entropy is used as the evaluation index to evaluate the learning regularity of the learner, the smaller the actual entropy is, the stronger the learning interval is, i.e., the stronger the learning time regularity is.

2.3 Similar Behavior Clustering and Analysis Based on Online Learning Behavior Data

2.3.1 Course-Specific Learning Behavior Clustering Methods

When students in a class study a certain course, different students have different learning habits, and ultimately the learning effect of the same lesson will be different. By clustering all students studying the same course and dividing them into different categories based on their learning behaviors, the relationship between the learning effect of each category of students and their learning behaviors can be found, which can help to help teachers to improve the quality of teaching.

As the data of students' learning process in each chapter is continuous, for example, each student enters the course at a different time and leaves the course at a different time, and each student learns at a different time in each chapter. Such data is difficult to model and analyze, so serialize the user's learning data. For each student, all the learning process click data of the student are first extracted, and these process data are sequenced in chronological order. The sequence of user online learning behavior is shown in Figure 1.

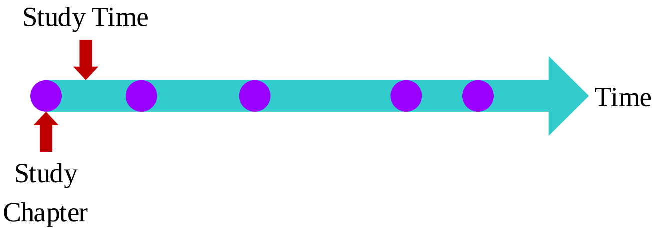

Figure 1: Sequence of online learning behaviors of users

This gives us the amount of time a student has spent studying a particular chapter. For example, a student learns 30 seconds in Chapter 1 and 1 minute in Chapter 2, we format the chapter as C_i, the learning time as T_i, and the learning time as T_i, which is calculated based on the start and end time durations of each chapter. Since the learning time is a continuous sequence, we discretize the time into 6 bucket intervals. Define the learning time mapped as bucket intervals as:

$$<1s,[1s,30s],(30s,1\min],(1\min,2\min],(2\min,5\min],>5\min \tag{7}$$

From this you can serialize the student learning process data into a string. For example a particular student learning formatting sequence is $C_1C_2C_3C_4C_5C_6C_7C_8C_9C_{10}C_{11}C_{12}C_{13}$. Each student's learning process is serialized into a string format.

For a particular course, similar learning behavior means that students learn the course chapters in a relatively consistent order and spend a similar amount of time on each chapter. Therefore, in this paper, we use the string substrings extracted from the serialized sequence of students' learning behaviors as features for comparison, and the distance between the strings of the sequence of learning behaviors is used to calculate the similarity of learning behaviors.

The learning behavior data is a string: $S=(s_1s_2s_3s_4s_5s_5\cdots s_n)$, where s_i is the i th element of the serialized learning data, which may be the chapter C_i also could be the learning time T, and n is the length of the string of formatted learning behaviors for that user. Define Q to be the set of all consecutive substrings of S as shown in equation (8).

$$Q(S)=\{(s_i,s_{i+1},s_{i+2},\cdots s_{i+k-1}),i\in[1,n+1-k],k\in[1,len(S)]\} \tag{8}$$

After calculating $Q(S)$ for each user, it is necessary to calculate the similarity distance between two user behaviors. For two user behavior strings S_1 and S_2, the probability r that all elements in $Q(S_1)$ appear in $Q(S_2)$ is calculated. If $r>0.5$, the two behavioral data are considered similar.

In order to achieve clustering among students, this paper uses a graph model approach to model the learning process data. First, each user is a vertex in the graph; then the user similarity is calculated, and if two users are similar, then these two users are connected by an undirected edge. Finally, a relationship graph between users is derived. We consider all mutually reachable students in the graph to be similar and mutually unreachable students to be dissimilar, so multiple student behavioral categories are derived by solving for the connected components in the relationship graph.

The results of modeling users' learning behaviors are shown in Fig. 2, where user 1, user 3, and user 4 have similar behaviors and form a connected subgraph, and user 2, user 5, and user 6 have similar behaviors and form another connected subgraph. The six users can be categorized into two classes based on connectivity. Before building a graph model, it is necessary to initialize the vertices and edges of a graph and then calculate the similarity of each user after serialization.

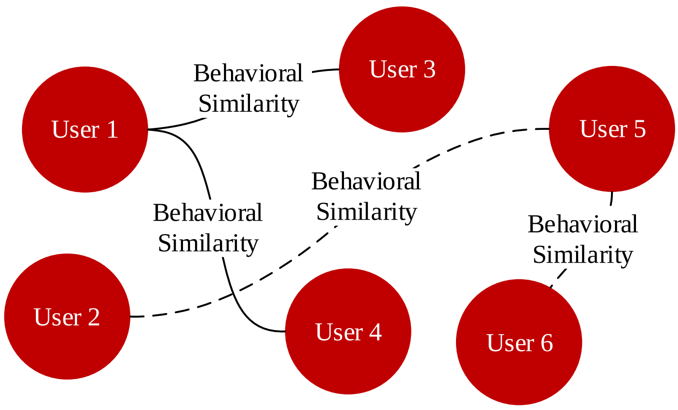

Figure 2: Modeling results for user learning behavior

2.3.2 Overall Learning Behavior Clustering Methods

Learning behaviors are highly correlated with the course itself because of the presence of course chapter-related factors in the modeling process. Considering that there are overall differences in student learning and that these differences are course-independent and holistic. Therefore, this section coarsens the model in 2.3.1 to analyze the overall differences in student learning.

Since students' online learning behavior habits are mainly expressed in the form of learning time, the coarse-grained learning behaviors in this section are similarly modeled in terms of learning time. For a single student, the data of all the learning of a particular student is extracted, and the number of data is defined as k; according to the category $T_1T_2T_3T_4T_5T_6$ divided by the mapping bucket of learning time, and the number of learning behavior entries of the learning time of the T_i category is denoted by X_i, the weight of the T_i category operation occurring in the j student R_{ji} is:

$$R_{ji} = \frac{X_i}{k}\left(i = 1, 2, 3, 4, 5, 6\right) \tag{9}$$

From this a vector of learning behaviors can be obtained for each student $v_j = \left(R_{j1}, R_{j2}, R_{j3}, R_{j4}, R_{j5}, R_{j6}\right)\left(j = 1 \cdots \text{ Total number of students}\right)$. The similarity between v_j is calculated to get the similarity of students' learning behavior habits. Define the similarity of two student behavior vectors as shown in equation (10).

$$\cos\left(v_i, v_j\right) = \frac{\sum_{i=1}^{6}\left(v_i \times v_j\right)}{\sqrt{\sum_{i=1}^{6}\left(v_i\right)^2} \times \sqrt{\sum_{i=1}^{6}\left(v_j\right)^2}} \tag{10}$$

The cosine similarity is used to define the similarity between different learning behaviors. The closer the cosine value is to 1, the more similar the two behaviors are; the closer the cosine value is to -1, the more different the two behaviors are. Based on the learning behaviors of all students, the behavior vector of each student is calculated, and then the similarity between two students is calculated, and it is considered that the two students with $0 < \cos\left(v_i, v_j\right) \leq 1$ are similar in their behaviors.

After calculating the two-by-two similarity between the students in terms of overall learning behaviors, the students are further modeled as a graph. Again, all students are treated as vertices in the graph and edges are connected based on the overall learning behavior similarity.

2.3.3 Learning state analysis based on behavioral clustering

For the clustering results, the learning behavior habits of different categories of students can be correlated and analyzed with the learning process data and learning effect data, and feedback to teachers. Teachers can use the feedback information to improve the teaching program and educate different categories of students in different directions.

The following statistical analyses can be made for students of different behavioral categories:

(1) Score distribution of students of different behavioral categories

The score is the most intuitive reflection of the learning effect of students. By counting the scores of the works submitted by students after learning, we can understand the mastery of the course. And then through the online learning behavior model for student clustering, you can score statistics for each category of students, analyze the different behaviors of each cluster of students, resulting in differences in scores, to understand the most suitable learning methods for students.

(2) Distribution of study time of students in different behavioral categories

The relationship between the total time students spend studying a particular course and the final score of the submitted work can be extracted from the learning behavior data. The total time spent studying a course is obtained by summing up the time students spent studying each section, and the work score is obtained by the course associated with the student's submitted work. By counting the relationship between the total study time and the score, it can be concluded whether the final work score will be higher if the study time is longer. At the same time, the average study time spent by a student to submit a qualified work for a course can be taken as the effective study time of the course, which is set as the recommended study time for the course, which can help students and teachers to check the learning status of the course.

3 Empirical Analysis of Learners' Behavioral Pattern Mining in Online Programming Courses

This paper collects data on the online learning activities of full-time undergraduate students from the online Learning Management System (LMS) at University A, as well as data on the online learning of in-service graduate students from University B's LMS. The undergraduate program consisted of 300 traditional full-time students with 266,863 behavioral logs. The graduate program consisted of 300 active students with a total of 251,646 behavioral log records, and the behavioral log data contained student IDs, IP addresses, timestamps, event types, and device information. In addition, student demographics and final grades were retrieved from each institution's data warehouse. In order to investigate the possible differences in the spatio-temporal behaviors of different types of students, this paper explores the spatio-temporal distribution patterns of online learning behaviors using 600 students as the study population. For the dynamics and cluster analysis part, only 300 traditional full-time students' behavioral data were used as samples for the experimental investigation in order to eliminate the influence of grade level, space-time, and other factors.

3.1 Analysis of the temporal and spatial distribution pattern of online learning behavior

3.1.1 Pattern recognition of different spatio-temporal behaviors based on entropy analysis

A higher entropy value usually has two possible meanings. For example, a higher temporal entropy value usually indicates that a student has studied in a higher number of study periods (i.e., more time periods) or that the distribution of study periods is more skewed (i.e., it is preferred to study in specific time periods). The scatter distributions of time entropy and location entropy for working students and full-time students are shown in Figure 3. The median values of time entropy and location entropy for working students are 2.11 and 1.01, respectively, while the median values of time entropy and location entropy for full-time students are 1.70 and 0.98, respectively.Using these median values, each category of learners can be divided into four subclusters with different time and spatial attributes. After subgrouping, each category of students has the following four subgroups: high time entropy-high location entropy group (HT-HL), high time entropy-low location entropy group (HT-LL), low time entropy-high location entropy group (LT-HL), and low time entropy-low location entropy group (LT-LL).

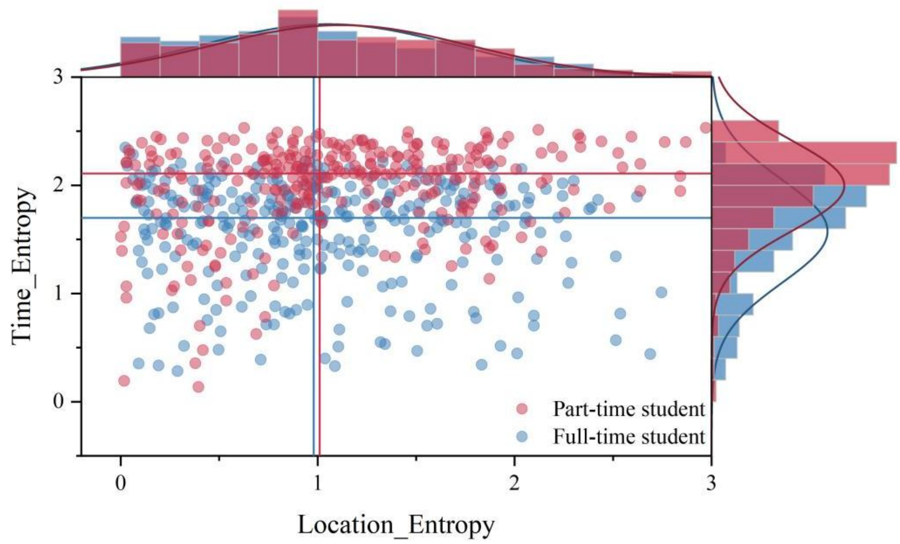

Figure 3: Scatter distribution of time entropy and position entropy

3.1.2 Characterization of groups with different temporal and spatial behavior patterns

(1) Comparison of temporal patterns among groups with different temporal behavior patterns

The results of temporal pattern comparison between different groups are shown in Table 1. First of all, the average total behavioral frequency of students in the high temporal entropy groups (i.e., HT-HL and HT-LL) is higher, reaching 701.37/859.24 and 425.64/356.87 among working students and full-time students, respectively.For full-time students, the study activities often occur in the evening study period, accounting for more than 45% of the time. For working students, students in the high time entropy group invested more time in weekday afternoons and weekend afternoons, while students in the low time entropy group invested more study time in weekday evenings. In addition, the time allocation of students in the high time entropy group is somewhat more even than that of students in the low time entropy group because they also allocate more study time on weekend mornings. Overall, there is no specific time pattern that leads to better or worse learning outcomes.

Table 1: Results of temporal pattern comparison between different groups

	Part-time student				Full-time student			
	LT-LL	LT-HL	HT-LL	HT-HL	LT-LL	LT-HL	HT-LL	HT-HL
Weekday morning	15.7%	17.3%	15.7%	18.3%	8.7%	9.7%	10.7%	14.3%
Weekday afternoon	27.3%	22.3%	26.3%	28.3%	24.3%	27.7%	22.3%	24.7%
Weekday evening	29.3%	27.7%	20.3%	17.3%	50.7%	46.3%	38.7%	34.3%
Weekend morning	3.7%	4.3%	10.3%	9.7%	1.7%	2.7%	6.7%	5.7%
Weekend afternoon	13.7%	12.7%	19.7%	19.7%	6.3%	6.3%	11.3%	10.3%
Weekend evening	10.3%	15.7%	7.7%	6.7%	8.3%	7.3%	10.3%	10.7%
Weekday	72.3%	67.3%	62.3%	63.9%	83.7%	83.7%	71.7%	73.3%
Weekend	27.7%	32.7%	37.7%	36.1%	16.3%	16.3%	28.3%	26.7%
Morning	19.4%	21.6%	26.0%	28.0%	10.4%	12.4%	17.4%	20.0%
Afternoon	41.0%	35.0%	46.0%	48.0%	30.6%	34.0%	33.6%	35.0%
Evening	39.6%	43.4%	28.0%	24.0%	59.0%	53.6%	49.0%	45.0%
Total behavior frequency	498.33	433.52	859.24	701.37	206.45	311.52	356.87	425.64
Pass rate	92.67%	87.33%	97.33%	93.67%	84.67%	78.33%	82.33%	76.33%

(2) Comparison of spatial patterns of groups with different temporal and spatial behavior patterns

After identifying the time periods during which different groups most often conduct their studies, it is necessary to further analyze the most common spatial patterns of study for each group. The study locations of full-time students were identified by back-parsing the IP addresses or GPS information from their behavioral logs, and the location was relatively accurate because the IP addresses were managed by the IT office on campus. However, for working students, their location was estimated by combining the time period and IP address. Since afternoon and night hours are the main study hours for both working and full-time students, this paper shows the top two study locations by location tag frequency for weekday afternoon, weekend afternoon, and night hours, and the results of comparing the top 2 study locations between different groups are shown in Table 2. For the two low positional entropy groups of working students (i.e., HT-LL and LT-LL), they mainly carried out their studies in the afternoons at their workplaces and in the evenings when they returned home, with percentages of 71.3%/72.3% and 62.7%/66.3%, respectively. The HT-LL population showed slightly greater positional variability than the LT-LL population when compared to the HT-LL and LT-LL populations. However, for the high location entropy groups (i.e., HT-HL and LT-LL), the frequency of “other” labels was higher than that of home or workplace during most study periods (because of the difficulty in distinguishing more precise locations by IP address), and thus it can be inferred that the primary study location of students in the high location entropy groups may still be home or workplace. It can be inferred that students in the high location entropy group may still study primarily at home or work. However, this percentage is lower than any other group. A similar pattern can be found among full-time students. However, for full-time students, their top two study locations are the same regardless of group. The results in Table 2 indicate that the top two study locations are classroom/library on weekday afternoons, classroom/dormitory on weekend afternoons, and dormitory/classroom in the evenings. However, the differences for the high location entropy group are much greater than for the low location entropy group.

Table 2: Comparison results of the top 2 learning positions between different groups

	Part-time student		
	Weekday afternoon	Weekend afternoon	Evening
LT-LL	Workplace/Other (72.3%/27.7%)	Home/Other (66.3%/33.7%)	Home/Other (78.3%/21.7%)
HT-LL	Workplace/Other (71.3%/28.7%)	Home/Other (62.7%/37.3%)	Home/Other (65.0%/35.0%)
LT-HL	Other/Workplace (56.0%/44.0%)	Other/Home (64.3%/35.7%)	Other/Home (65.7%/34.3%)
HT-HL	Other/Workplace (66.3%/33.7%)	Other/Home (68.7%/31.3%)	Other/Home (76.0%/24.0%)
	Full-time student		
	Weekday afternoon	Weekend afternoon	Evening
LT-LL	Classroom/Library (70.3%/19.7%)	Classroom/Dormitory (73.3%/8.7%)	Dormitory/Classroom (66.7%/23.3%)
HT-LL	Classroom/Library (57.3%/32.3%)	Classroom/Dormitory (72.7%/19.7%)	Dormitory/Classroom (52.7%/30.3%)
LT-HL	Classroom/Library (36.0%/32.7%)	Classroom/Dormitory (39.7%/21.3%)	Dormitory/Classroom (27.3%/30.0%)
HT-HL	Classroom/Library (19.7%/36.0%)	Classroom/Dormitory (34.3%/38.7%)	Dormitory/Classroom (14.7%/26.3%)

3.2 Dynamics Analysis of Online Learning Behavior

Behavior obeys the power law distribution of the driving model can be summarized into two categories, one belongs to the task-driven, this type of behavior obeys the power law distribution is because people tend to prejudge the priority of the task, when the task is urgent will be given priority to complete, and vice versa, the task is not urgent, may not be processed immediately, resulting in the completion of the task of the behavior of the time distribution of the power law characteristics of the task a large portion of the task can be processed very quickly, the waiting time is short, and a A small number of tasks have a long waiting time. Another type of drive model where behavior follows a power law distribution is called the interest-driven model. Some of the behaviors still have interval distributions that follow a power law distribution, however, these behaviors do not involve the concept of task and therefore are not applicable to the task-driven model. Theoretically, on the one hand, online resource-based learning emphasizes student-oriented learning, and students have a high degree of freedom in order to be more susceptible to changes in personal interests. However, on the other hand, in online teaching, teachers will issue course learning supervision announcements from time to time, and the participation rate of online courses will affect the final grade of the course, therefore, this paper speculates that the learning behavior should also be influenced by teachers. In this regard, this paper intends to further mine the learner-resource interaction behavior and explore the learner-resource interaction behavior law.

3.2.1 Statistical characteristics of group behavior

An empirical analysis of time intervals of learner-resource interaction behaviors at the group level was conducted to investigate whether the time intervals of learner-resource interaction behaviors obeyed a power law distribution. Due to the large amount of group behavior data, the distribution can be fitted directly using the least squares toolkit on MATLAB. The results of the time interval distribution of learner-resource interaction behaviors are shown in Figure 4, where the horizontal coordinates are the time intervals and the vertical coordinates are the corresponding cumulative distributions. The learner-resource interaction behavior at the group level indeed obeys a power law distribution with a power index α of 1.913 and R^2 of 0.987, which is a good fit. Combined with the actual analysis of the course, when learners are learning online resources, the initial occurrence of the behavior may be due to the weekend time is generous, the offline classroom on Thursday assigned homework or the new knowledge of the professor inspired the desire to know, etc., resulting in the peak of students' interest in learning online resources, with the passage of time, the interest in learning is gradually diminishing, and a gradual reduction of the frequency of resource access can be captured in the behavior.

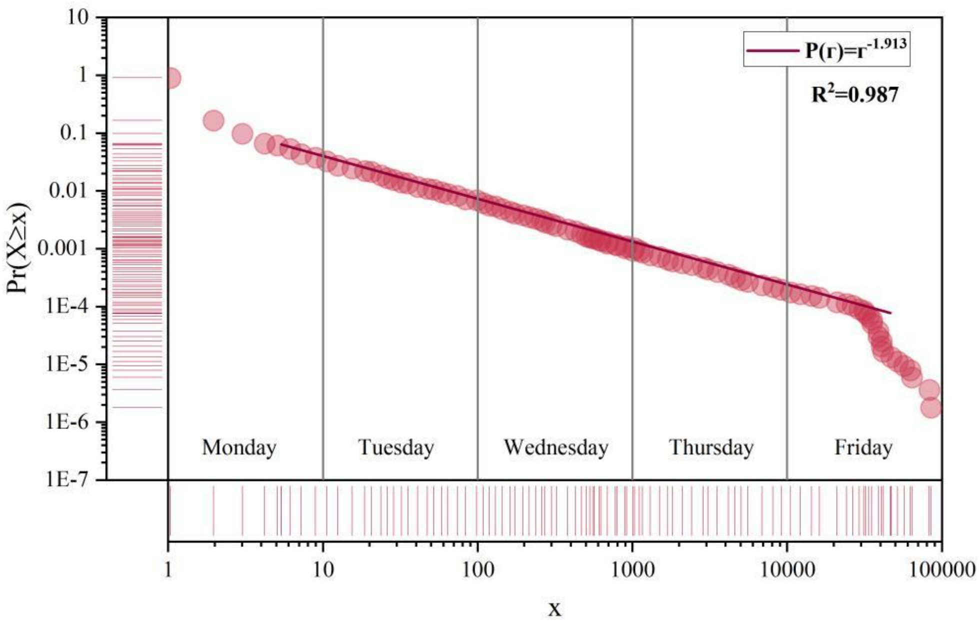

Figure 4: Results of time interval distribution of interaction behaviors

3.2.2 Statistical Characteristics of Individual Behavior

In order to explore whether the time interval distribution of resource interaction behaviors at the individual level also obeys a power law distribution, the behavioral sequences with the same user ID number are filtered out as the resource access time sequences of individuals, which are used to analyze the time interval distribution of behaviors at the individual level. In this study, the resource access time sequence of one student was randomly selected from each of the URL numbers of 1000-1500, 1500-2000, 2500-3000, and 3500-4000, and the abruptness of the behavior is more obvious in the individual than in the group, which will not be repeated here. In the study of learner-resource interaction behavior at the individual level, this paper adopts the great likelihood estimation (MLE), which resists the tail fluctuation problem brought about by finite samples by calculating the complementary cumulative distribution function of the samples, so as to verify whether the observed data obeys a power law distribution.The distribution of the time intervals of resource interaction behavior of the four learners (No. S1~S4) is shown in Fig. 5, and the analysis reveals that the investigated individual resource access behaviors still obey the power law distribution, and the power indices are all around 1.9. This suggests that for individual learners, usually after a long pause, learners will access online resources again with high frequency, and the learner-resource interaction behavior at the individual level is also a kind of bout behavior.

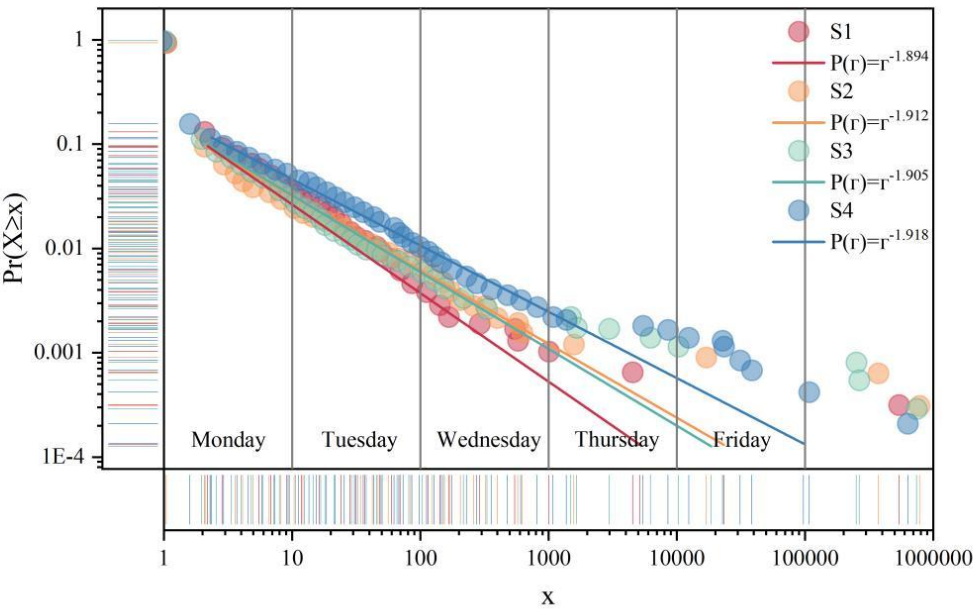

Figure 5: Resource interaction behavior time interval distribution of four learners

3.2.3 Resource Interaction Behavior Dynamics Model Assumptions and Simulation

After the above analysis, the learner-resource interaction behavior will show periodicity within the threshold, but the teacher's online announcement posting behavior will affect the learners' resource access frequency, therefore, the interest decay function with only univariate parameters can not describe the learner-resource interaction behavior well, and it is necessary to introduce announcement posting influence factor. As a result, we can use the above behavioral analysis law to make a kinetic model of learner-resource interaction behavior applicable to the blended teaching mode. First, the following assumptions are made about the model:

(1) Time is discrete, and the smallest unit of time step is 1. The probability of learners accessing the platform resources is utilized to express their learning interest in the online resources of blended teaching, and the maximum value of interest is set to 1.

(2) Since the learners' behavior of accessing the platform resources shows a cyclical change rule, a threshold is set T. In a cycle, with the passage of time, the learners' interest in learning the resources decreases from the maximum value, and eventually, after the end of a cycle, the interest will return to the initial maximum value.

(3) If the teacher posts an online announcement, it will stimulate students' interest in resource learning again to some extent. In other words, the teacher's announcement posting behavior will reduce the degree of decay of learners' interest in resource learning to a certain extent, but it will not affect the entire length of the behavioral cycle, and the parameter λ is used to denote the influence factor of the teacher's announcement posting, which regulates the magnitude of the degree of decay.

In this paper, based on the linear decay of participants' interest over time, we added the announcement release influence λ and fixed the interest cycle decay degree β to 0.8 to verify the validity of the empirical data. Then the announcement release influence factor λ was taken as 0 and 0.75 respectively to simulate the model, and a comparison of the fitted curves of the simulation results is shown in Figure 6. When $\beta = 0.8$, $\lambda = 0$, it means that the learner-resource interaction behavior is not affected by the teacher's announcement posting behavior, at this time, the power index is 1.984, and the power index produced under the value of this parameter is 1.9828, which is more in line with that obtained from the derivation of the formula. When $\beta = 0.8$, $\lambda = 0.75$, that is, after the introduction of the teacher's announcement release influence factor, the power index is 2.5068, and the formula derived from the power index of 2.5 is also more consistent with the model has good explanatory power. It can be found that in the case of no announcement influence factor, the students' behavior is not affected by the teacher's announcement release, and their interest naturally declines, and the slope of the corresponding power index is significantly smaller than that in the case of having the influence of the teacher's announcement release, which confirms that the behavior of the teacher's announcement release positively affects the learner-resource interaction behavior.

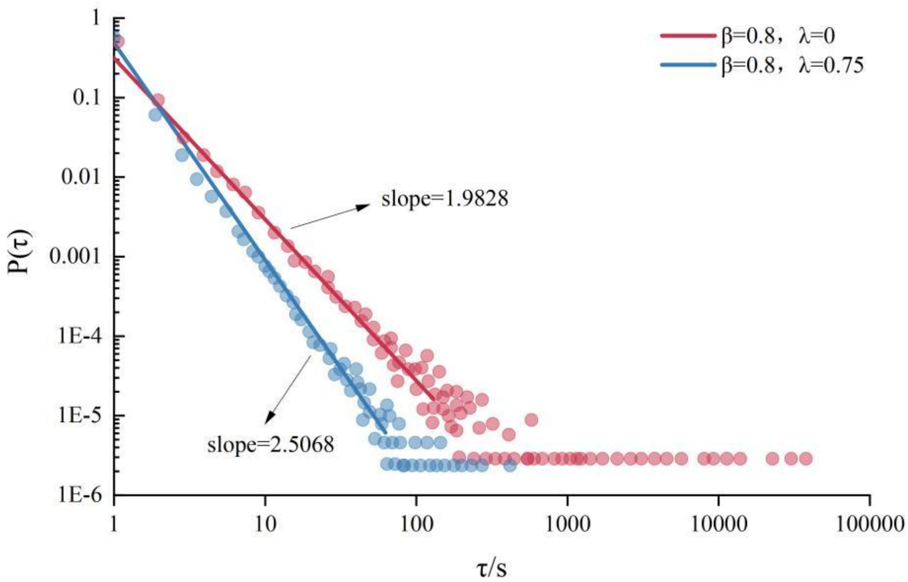

Figure 6: Numerical simulation results of the kinetic model

3.3 Learning Situation Analysis for Online Learning Behavior Clustering

3.3.1 Cluster Analysis of Online Learning Behaviors

According to the clustering results output from the k-means algorithm, the clustering results of the online learning behavior data are shown in Figure 7 by visualizing the data. The clustering results show three different clusters with high intra-cluster similarity and low inter-cluster similarity. The three clusters generated from the clustering results of the experimental data show the online learning behavior preferences of different groups, and by analyzing the online learning behavior data in the same clusters, it explores whether there is a certain relationship between the online learning behaviors of the same type of learners and their learning performance.

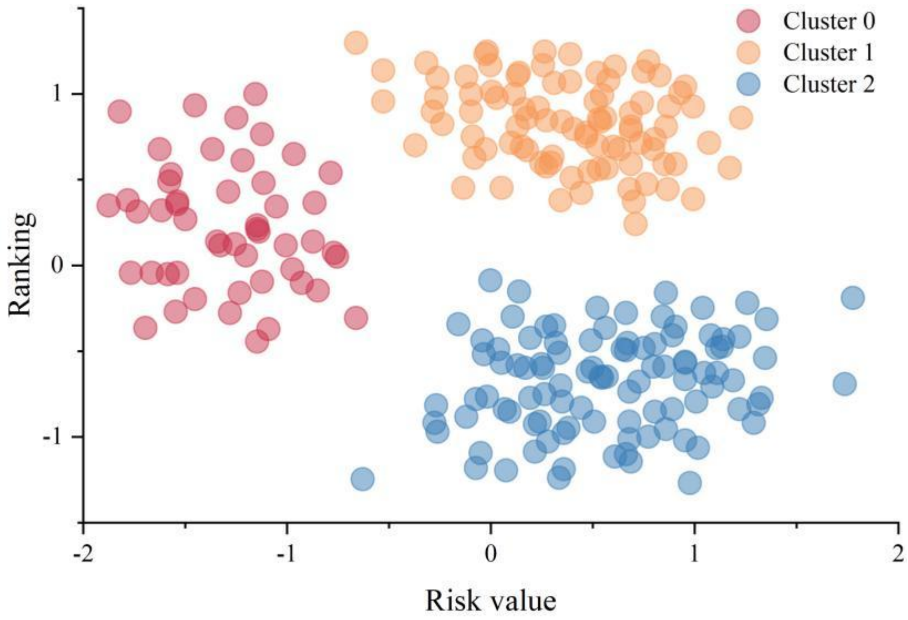

Figure 7: Clustering results of online learning behavior data

Further hierarchical clustering was performed to obtain a hierarchically organized student profile as shown in Fig. 8. Each instance on the x-axis represents a student, and the horizontal line on the y-axis represents the Euclidean distance between instances. The students were clustered into three categories, denoted as S1, S2, and S3.

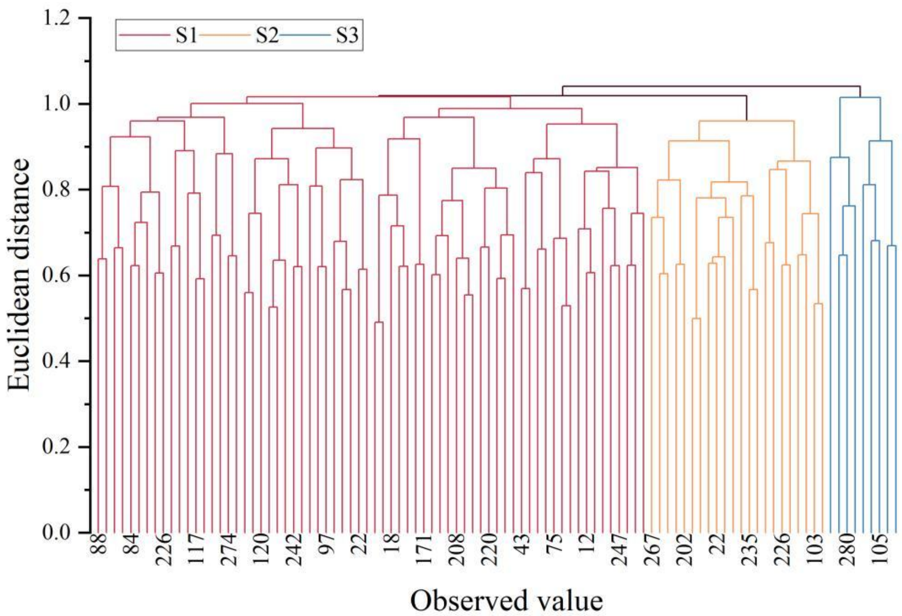

Figure 8: Stratified organization of students

The students in these three categories were then integrated and named Highly Engaged Students, Randomly Engaged Students and Lowly Engaged Students based on their online learning behaviors, and the three metrics in the learner profile, code quality, lab scores, and final quiz scores, were selected for statistical purposes. The code quality and lab scores are normalized values, which are multiplied by 100 for clearer observation in the presentation. After clustering, the performance data of these three types of students are shown in Figure 9, from which it can be seen that the highly engaged students have better final quiz scores, with a mean value of 88.03, and also have leading scores in usual experiments and code quality. Randomized students had a more average final quiz score, with a mean of 76.07, and their usual lab scores were in the middle of the pack. The low-involvement students' usual lab scores and final quiz scores were less favorable, with mean values below 75. On the indicator of code quality, there was almost no gap between the randomly involved and low-involvement students. This shows that the students who achieved excellent scores on the final programming quiz had already highlighted certain strengths in their usual programming experiments.

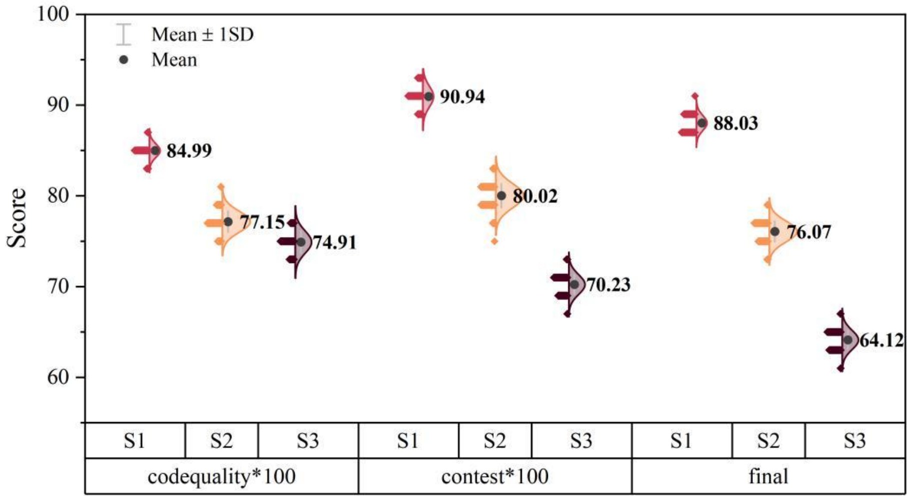

Figure 9: Performance data for the three categories of students

3.3.2 Learning situation analysis based on learner profiling

The distribution of correct rates for each question type in all experiments shows the overall mastery of the students for each knowledge point. The results of students' correct rates for each question type are shown in Figure 10. The questions in the experiments were categorized into four main types: basic, branching, loops, arrays and functions, with basic questions being simple input/output or simple arithmetic questions. It can be seen that the students have the lowest level of mastery for the cyclic question types, with an average correct rate of only 50.17%. For branching questions, the students' mastery is better, with an average correct rate of 70.4%. In the subsequent teaching process, it is necessary to focus on strengthening the practice and guidance of the circular chapter.

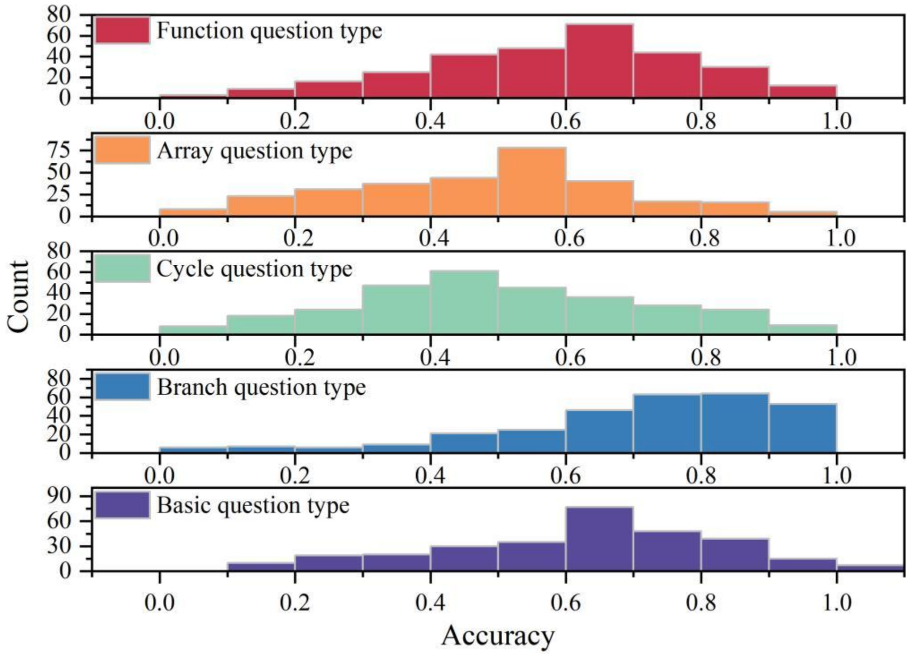

Figure 10: Results of students' correct rate of each question type

4 Behavioral Intervention Strategies for Learners in Online Programming Courses

One of the important goals of educational data mining and learning analytics is to provide teachers with feasible suggestions. This paper analyzes the learning behaviors of different types of learners in online programming environments, which objectively and clearly reacts to the learners' learning status. In order to further improve students' learning ability in online programming environment, this paper proposes specific intervention strategies from the perspective of teachers based on the patterns and problems found in the previous paper.

First, since “on-the-go” learning styles are negatively correlated with learning outcomes, instructors can design more efficient online courses to encourage students to achieve consistent and stable levels of engagement in order to maximize student success. Students who are required to travel frequently (i.e., have a high degree of location transitions) may be advised to take fewer courses per semester to reduce their course load and successfully complete their academic tasks each semester.

Second, depending on the cycle instructors can post online announcements at low weekly predicted access values. Since the content of the offline course is fixed, the content of the announcements can include task completion reminders and resource recommendations related to the knowledge points studied offline each week, so as to achieve the goal of allowing students to maintain the continuity of their study time and to adopt the habit of studying in small steps. For example, in the first chapter of learning, the teacher can set the recommended content as a reminder of the completion of course tasks on the Friday when the number of visits decays to the lowest, which is not only released in the bulletin area of the course platform, but also sent to the students' mailboxes associated with the platform.

Third, different types of learners present different online learning behavior preferences, and teachers should focus on individual differences of learners. They should provide appropriate learning support for learners who do not use different types of learning and intervene in the learning process to optimize their learning paths. Teachers should find out from the group portrait how students perform in the programming learning process and how well they master each type of knowledge, so as to develop a more macro teaching plan. Individual portraits of students will reveal what each student needs to improve in the programming learning process, as well as the ups and downs of the learning process, so that more targeted guidance can be provided.

5 Conclusion

This paper successfully analyzes different types of online learning behaviors by constructing a behavioral analysis framework for online programming learning, using Universities A and B as study cases. Through the behavioral analysis to find out the behavioral patterns, from the dimension of course instructors, further stakeholder intervention suggestions are proposed to help teachers optimize the design of online teaching activities, and the main conclusions are the following three points.

(1) Working students present high average time entropy and location entropy values, and working students are more inclined to take advantage of online learning anytime and anywhere compared to full-time students. In addition, the difference between the high time entropy group and the low time entropy group of working students in terms of the time period of study is smaller. Students in the low positional entropy group have better learning outcomes than those in the high positional entropy group, and high temporal entropy does not necessarily lead to good learning outcomes because learning outcomes are also affected by spatial learning patterns.

(2) The universality of the “power law presentation” of learner-resource interaction behavior is confirmed at the group and individual levels, in which the group behavior obeys a power law distribution with a power index of 1.913, and the power index of the individual behavior is around 1.9.

(3) According to the online learning behaviors, students are clustered into three categories: high-involvement students, random participation students and low-involvement students, and it is found that the final quiz scores are positively correlated with the usual experimental scores, and the students who have better scores on the final quiz also maintain the excellent level of the usual experimental scores. Among the five question types, students had the lowest level of mastery for the cyclic question type and the best level of mastery for the branching question type.

Funding

This work was supported by the Research Funding Project for Introduced Talents of Southwest University of Political Science and Law (2023-XZRCXM022).

About the Author

Yang He was born in Nanchang City, China in 1991. He obtained his master's degree from Shaanxi Normal University and currently works at the School of Mathematics and Physics of Fuzhou Preschool Education College. His main research direction is modern educational technology.

Dun Mao was born in Fuzhou City, Jiangxi Province, China in 1986. He graduated from Zhengzhou University of Light Industry with a bachelor's degree. He is currently working at the School of Mathematics and Physics, Fuzhou Preschool Education College, and his main research focus is computer application technology.

Jiayu Li was born in Jiangxi, P.R. China, in 1995. She obtained a bachelor's degree from Southwest University. I am currently working at Fuzhou Preschool Education College. My main research directions is modern educational technology.

References

[1] Chavali, A., & Biradar, V. (2022). Digitalization of education—challenges and measures for promotion. In Evolution of Digitized Societies Through Advanced Technologies (pp. 33-44). Singapore: Springer Nature Singapore.

[2] Alencar, A. R. D., Castro, A. M., Soares, D. T. D., Amaral, L. D. S. V. E. V., & Gomes, R. L. R. (2024). Digitalization and participation in education: Opportunities, challenges, and initiatives in the digital world. IOSR Journal of Humanities and Social Science, 29(10), 18-26.

[3] Jie, W., Hai-yan, L., Biao, C., & Yuan, Z. (2017, June). Application of educational data mining on analysis of students' online learning behavior. In 2017 2nd International Conference on Image, Vision and Computing (ICIVC) (pp. 1011-1015). IEEE.

[4] Miller, L. D., Soh, L. K., Samal, A., Kupzyk, K., & Nugent, G. (2015). A Comparison of Educational Statistics and Data Mining Approaches to Identify Characteristics That Impact Online Learning. Journal of Educational Data Mining, 7(3), 117-150.

[5] Kimmons, R., & Veletsianos, G. (2018). Public internet data mining methods in instructional design, educational technology, and online learning research. TechTrends, 62(5), 492-500.

[6] Agbo, F. J., Oyelere, S. S., Suhonen, J., & Adewumi, S. (2019, November). A systematic review of computational thinking approach for programming education in higher education institutions. In Proceedings of the 19th Koli calling international conference on computing education research (pp. 1-10).

[7] Lin, H. C. K., Tseng, C. H., & Chen, N. S. (2025). Enhancing programming education. Educational Technology & Society, 28(2), 279-294.

[8] Lindberg, R. S., Laine, T. H., & Haaranen, L. (2019). Gamifying programming education in K-12: A review of programming curricula in seven countries and programming games. British journal of educational technology, 50(4), 1979-1995.

[9] Williamson, B., Bergviken Rensfeldt, A., Player-Koro, C., & Selwyn, N. (2019). Education recoded: policy mobilities in the international 'learning to code'agenda. Journal of Education Policy, 34(5), 705-725.

[10] Moubayed, A., Injadat, M., Nassif, A. B., Lutfiyya, H., & Shami, A. (2018). E-learning: Challenges and research opportunities using machine learning & data analytics. IEEE Access, 6, 39117-39138.

[11] Du Plessis, M. C., Niu, G., & Sugiyama, M. (2014). Analysis of learning from positive and unlabeled data. Advances in neural information processing systems, 27.

[12] Mehenaoui, Z., Lafifi, Y., & Zemmouri, L. (2022). Learning behavior analysis to identify learner's learning style based on machine learning techniques. Journal of Universal Computer Science, 28(11), 1193.

[13] Chengjiu, Y. I. N., Uosaki, N., Hui-Chun, C. H. U., HWANG, G. J. G. J., Gi-Zen, L. I. U., HWANG, J. J., ... & Tabata, Y. (2017, December). Learning behavioral pattern analysis based on students' logs in reading digital books. In International Conference on Computers in Education.

[14] Ye, Z., Jiang, L., Li, Y., Wang, Z., Zhang, G., & Chen, H. (2022). Analysis of differences in self-regulated learning behavior patterns of online learners. Electronics, 11(23), 4013.

[15] Yoon, M., Lee, J., & Jo, I. H. (2021). Video learning analytics: Investigating behavioral patterns and learner clusters in video-based online learning. The internet and higher education, 50, 100806.

[16] Kinnebrew, J. S., & Biswas, G. (2012). Identifying Learning Behaviors by Contextualizing Differential Sequence Mining with Action Features and Performance Evolution. International Educational Data Mining Society.

[17] Mojarad, S., Essa, A., Mojarad, S., & Baker, R. S. (2018, May). Data-driven learner profiling based on clustering student behaviors: learning consistency, pace and effort. In International conference on intelligent tutoring systems (pp. 130-139). Cham: Springer International Publishing.

[18] Emara, M., Rajendran, R., Biswas, G., Okasha, M., & Elbanna, A. A. (2018). Do Students' Learning Behaviors Differ when they Collaborate in Open-Ended Learning Environments?. Proceedings of the ACM on human-computer interaction, 2(CSCW), 1-19.

[19] Shelle, G., Earnesty, D., Pilkenton, A., & Powell, E. (2018). Adaptive learning: An innovative method for online teaching and learning. The Journal of Extension, 56(5), 17.

[20] Tsai, C. W. (2013). An effective online teaching method: The combination of collaborative learning with initiation and self-regulation learning with feedback. Behaviour & Information Technology, 32(7), 712-723.

[21] Real, E. M., Pimentel, E. P., & Braga, J. C. (2021, October). Analysis of learning behavior in a programming course using process mining and sequential pattern mining. In 2021 IEEE Frontiers in Education Conference (FIE) (pp. 1-9). IEEE.

[22] Pereira, F. D., Oliveira, E. H., Oliveira, D. B., Cristea, A. I., Carvalho, L. S., Fonseca, S. C., ... & Isotani, S. (2020). Using learning analytics in the Amazonas: understanding students' behaviour in introductory programming. British journal of educational technology, 51(4), 955-972.

[23] Piech, C., Sahami, M., Koller, D., Cooper, S., & Blikstein, P. (2012, February). Modeling how students learn to program. In Proceedings of the 43rd ACM technical symposium on Computer Science Education (pp. 153-160).

[24] Carter, A. S., & Hundhausen, C. D. (2017, March). Using programming process data to

detect differences in students' patterns of programming. In Proceedings of the 2017 ACM SIGCSE Technical Symposium on Computer Science Education (pp. 105-110).

[25] Rivers, K., Harpstead, E., & Koedinger, K. R. (2016, September). Learning curve analysis for programming: Which concepts do students struggle with?. In ICER (Vol. 16, pp. 143-151).

[26] Wang, W. (2025, June). Analysis of Online Learning Behavior in Java Programming Design Courses using a Hybrid Machine Learning Approach. In 2025 International Conference on Intelligent Computing and Knowledge Extraction (ICICKE) (pp. 1-7). IEEE.

[27] Hoq, M., Brusilovsky, P., & Akram, B. (2023, July). Analysis of an explainable student performance prediction model in an introductory programming course. In the 16th International Conference on Educational Data Mining (EDM 2023).

[28] López-Pernas, S., Saqr, M., & Viberg, O. (2021). Putting it all together: Combining learning analytics methods and data sources to understand students’ approaches to learning programming. Sustainability, 13(9), 4825.

[29] Filvà, D. A., Forment, M. A., García-Peñalvo, F. J., Escudero, D. F., & Casañ, M. J. (2019). Clickstream for learning analytics to assess students’ behavior with Scratch. Future Generation Computer Systems, 93, 673-686.

[30] Falloon, G. (2016). An analysis of young students' thinking when completing basic coding tasks using Scratch Jnr. On the iPad. Journal of Computer Assisted Learning, 32(6), 576-593.

[31] Song, D., Hong, H., & Oh, E. Y. (2021). Applying computational analysis of novice learners' computer programming patterns to reveal self-regulated learning, computational thinking, and learning performance. Computers in human behavior, 120, 106746.

[32] Oeda, S., & Hashimoto, G. (2017). Log-data clustering analysis for dropout prediction in beginner programming classes. Procedia computer science, 112, 614-621.