IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received May 22, 2022, accepted June 13, 2022, date of publication June 17, 2022, date of current version June 23, 2022.
Digital Object Identifier 10.1109/ACCESS.2022.3184111

Exploration and Visualization of Learning Behavior Patterns From the Perspective of Educational Process Mining

GUIYUN FENG, MUWEI FAN, (Member, IEEE), AND CHUCHU AO
School of Management, Guizhou University, Guiyang 550025, China
Corresponding author: Muwei Fan (mwfan@gzu.edu.cn)

This work was supported in part by the Construction Project of 2017 First-Category Discipline in Guizhou: Big Data Discipline Group under Grant GNYL[2017]005.

ABSTRACT The data in modern educational information systems are not given enough attention and are not fully utilized. Therefore, the motivations of our study are to preliminarily explore learning behavior patterns by applying process mining to educational datasets, and construct prediction models based on previous learning behavior. The data in modern educational information systems can be used by teaching managers to analyze various aspects of the educational process from different perspectives. The study chooses three datasets randomly which include student number, courses and grades attributes from a university's educational information systems. This paper firstly applies system clustering to give an overview of students' academic performance, and roughly determines clustering number. In order to ensure the accuracy which is relevant to the analysis of students' learning behavior patterns, a semi-biased statistic is proposed to quantitatively determine clustering number. Then, the data are clustered by fast clustering algorithm, and the clustering effect is cross-validated which is aimed at accurately analyzing the behavior patterns of student groups and using data visualization technology to visualize different student groups. Finally, the support vector machine is used to construct a classifier for predicting the learning behavior pattern, and the parameters in the support vector machine are optimized by Bayesian optimization, genetic algorithm optimization and whale optimization algorithm respectively. The research revealed that 1) in the equal test of the group mean, when the significance of most courses is less than 0.05, it means that there is a significant difference among different categories. In this case, using the semi-biased statistic proposed in the paper is helpful to improve clustering effect. 2) The better the students learn, the better the clustering effect of the category which they belong to is. 3) Whale optimization algorithm performs the best.

INDEX TERMS Educational process mining, learning behavior patterns, student profiles, support vector machines, system clustering, the semi-biased statistic.

I. INTRODUCTION

Process mining (PM), as a branch of data mining (DM), puts forward process-oriented requirements for DM. The main idea of PM is to mine hidden processes from data, which provides a connection between data mining and process analysis. [1] PM is a growing and promising research field that focuses on understanding process and helping to obtain more important discovery during actual execution [2]. PM applies robust methodology including data mining and machine learning for pattern recognition [3]. Educational process mining (EPM) is the application of process mining technology in the field of education. It uses the recorded learning behavior data to explore the learning behavior pattern and predict the learning effect, which is conducive for teachers to understand the habits of students' learning process and find the factors affecting their performance. EPM is a new field in the discipline of educational data mining (EDM), which utilizes log data in educational information systems to discover, analyze, and improve educational process [4], [5]. The main purpose of EPM is to improve learning effect and innovate educational management, which can provide support for teaching decision-making.

EPM and EDM are all the applications of mining techniques in the education field, but EPM pays more attention to data mining in the educational process. Compared with EDM,

The associate editor coordinating the review of this manuscript and approving it for publication was Alba Amato.

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License.
For more information, see https://creativecommons.org/licenses/by-nc-nd/4.0/

there is less research on EPM. The educational process is centered on promoting students' development. The data in modern educational information systems are not given enough attention and are not fully utilized. If we want to improve the quality of educational process, we had better make full use of the data in the educational process. Therefore, how to mine and analyze students' learning behavior patterns through the educational process is an urgent problem to be solved. Based on this, three subdivided research questions in this paper are as follows:

1) How to explore the learning behavior patterns?
2) How to present the learning behavior patterns?
3) How to use the learning behavior patterns to innovate educational management?

In terms of the research questions, the main contributions of this paper are as follows:

1) The boundaries of EDM are expanded by paying more attention to the data in the educational process and studying learning behavior patterns from the perspective of EPM. Learning behavior patterns are mainly reflected in the academic performance. We can explore the learning behavior patterns by analyzing academic performance with the assistance of mining techniques.
2) Use data visualization technology to vividly present the learning behavior patterns. The introduced student profile is in line with the concept of higher education development, which has strong practical guiding significance. Only if the analyzed results are vividly presented can stakeholders utilize relevant information to improve the performance of students and make progress in the education field. We consider radar chart which reflects multi-index data as displaying the group course profiles of students.
3) Use the support vector machine that is suitable for small sample classification on the basis of optimizing parameters, and predict the future learning behavior pattern to achieve the goal of innovating educational management. Different from common application of machine learning techniques, we apply three optimization methods to make the parameters better to use in the classification algorithm so as to reduce prediction errors.

The remaining sections of this paper are organized as follows. Section II introduces the related work on the research topic in this paper and gives the research framework. Section III gives a brief description of the research methods used in this paper. Section IV presents the exploration, analysis, visualization, and prediction of learning behavior patterns. Section V summarizes the full text and looks forward to the potential research directions in this field.

II. RELATED WORK

A. PROCESS MINING

The basic idea of process mining was first proposed by Cook and Wolf in 1995 [6]. In 1998, Agrawal formally named it process mining [7]. The basic idea of process mining is to use log events to extract knowledge in complex information system to monitor and improve operational process. Traditional process analysis focuses on the establishment of models and does not make full use of the data in the running process. Data mining technology is data-centric, and the end-to-end process is unclear. The emergence of process mining solves the gap between model-based process analysis (such as simulation) and data mining, and is considered as an important bridge connecting data science and process science [8]. The three types of process mining are process discovery, conformance checking and process enhancement [2]. Each type focuses on different research topics. The applications in the research field originated from business [9]–[11], focusing on health [3], [12]–[15], information and communication [16], industry [17], education and other fields.

De Weerdt *et al.* [9] proposed an event log analysis framework based on process mining, demonstrated the usefulness and flexibility of process mining techniques, and conducted case studies to reveal inefficiencies in business organizations. Werner's model provides accurate information about control flow from an accounting perspective and is less complex than models using temporal correlations. Werner also evaluated the model using three real-world datasets [10]. Using machine learning techniques and considering various types of events, Borkowski *et al.* [11] demonstrated how to perform fault prediction in business processes, and evaluated the scenarios using two commercial datasets, showing that the proposed method is able to predict faults with high accuracy. Zayoud *et al.* [12] proposed a new procedural learning framework based on probabilistic learning and predicate logic, and applied it to medical big data, which can be used to predict the frequency and correlation of medical events. To identify opportunities in the process, Martin *et al.* [13] conducted a thorough process analysis that could be based on real process execution data captured by health information systems, and made recommendations to improve the usability and understandability of medical process mining. Martin *et al.* [14] proposed a new algorithm to detect parallel, sequential and concurrent batching of multiple connected tasks, and evaluated on logs generated by business process simulators as well as real logs obtained from hospital digital systems. De Roock and Martin [15] also expressed their views and opinions about process mining on the latest technologies in the medical field. Dallagassa *et al.* [3] summarized the opportunities and challenges faced by the application of process mining in the medical field. Diamantini *et al.* [16] used process mining techniques to analyze the movement patterns of social media users. Myers *et al.* [17] proposed a novel process mining anomaly detection method that uses industrial control system data logs and conformance checking to identify anomalous behaviors such as cyber attacks.

B. EDUCATIONAL PROCESS MINING

Educational process is a two-way activity process carried out by educators and educated people around educational goals.

EPM takes educational events as the research object, and realizes the extraction, analysis and visualization of educational process data [18]. Most of the research objects in the application of EPM are easily observed single channel behavior data, such as mouse click [19], test score [20], writing document [21] and course mining [22].

Trcka and Pechenizkiy [4] proposed a new framework for EPM that assumes that a set of pattern templates can be pre-defined for efficient centralized mining. They used the student database to verify through examples of course modeling, mining and conformance checking. Bannert *et al.* [23] showed how various methods developed in process mining research can be applied to identify process patterns in self-regulated learning events, and how process mining methods can be used to test self-regulated learning process models. Ariouat *et al.* [24] proposed a two steps approach of clustering to improve educational process mining. The first step consists of creating clusters based on employability indicators and the second step consists of clustering the obtained clusters based on traces profiles in order to refine the obtained results from the first step. It is found that this approach optimizes the performance of the obtained model. Maita *et al.* [25] conducted an exploratory study to use artificial neural networks to extract knowledge from an unstructured process in the distance learning domain. Results suggest that applying either classical process mining or modern data mining techniques would result in significant benefits for this domain. Maldonado-Mahauad *et al.* [26] addressed the problem of combining mining behavior patterns with a validated self-reporting tool, and used process mining to extract interaction sequences from tracking learner behavior in three massive open online courses. Using a process-oriented approach, Juhanak *et al.* [27] studied the use of process mining methods in the context of online learning. In order to explore various possible student behavior patterns, they also analyzed the interaction of students from different courses and different environments in online quiz behavior. Modi and Jagtap [28] proposed an algorithm of filtering tool which can recognize and block all non learning sites by matching the multiple patterns like text, video and images of the web pages by web content mining. Applying process mining techniques to take course trajectories as research process, Salazar-Fernandez *et al.* [29] analyzed the course trajectories of students based on courses they failed and validated the proposed model, finding that specific courses are associated with dropout rates. Akhuseyinoglu and Brusilovsky [30] utilized sequential pattern mining methods to construct individual models of learners' practical behaviors and explored connections between learner behaviors and incoming and outgoing parameters of the learning process. It is shown that data-driven individual difference models significantly outperform traditional individual difference models in predicting important parameters of the learning process, such as performance and engagement.

C. STUDENT PROFILE

User profile was proposed by Cooper in 2004, which refers to the target user model based on a series of real data [31]. Student profile is the application of user profile in the field of education, which can be used to identify and label student groups with different attributes. Clustering analysis is designed to identify students with different abilities to support the presentation and characterization of student profiles.

Trandafili *et al.* [32] conducted experiments on real datasets, and revealed the relationship between student profiles and course performance through clustering and association rule mining. Park *et al.* [33] proposed a student profile system using learner multidimensional feature analysis to extract learner feature information. The extracted learner feature information is automatically constructed into a personalized student profile through the learner description system. Dwivedi and Bharadwaj [34] proposed a student profile merging scheme based on learning style, knowledge level and learner category, and proposed a group recommendation collaborative method based on unified student profiles. Li *et al.* [35] extracted seven features from three aspects of students based on students' campus card consumption records, assigned three different levels of values to each feature to illustrate the consumption differences among students, and stuck labels on students to describe their campus consumption behavior, thereby constructing the student profile. Li and He [36] built a model with five dimensions of student information. Based on the methods of data collection, processing and mining, students' feature attributes were extracted and student profile was constructed.

Previous studies on EPM mainly involve in process discovery, conformance checking and process enhancement. Process discovery is embodied in the application of data mining technology to analyze and evaluate the educational process. Conformance checking behaves as the evaluation index to test the effect of data mining technology application. Process enhancement shows that deep-seated information and knowledge are obtained through the analysis results to improve the learning behavior patterns. Our study is dominated by the three stages and has some improvements which make our study different from others in three stages. The student number and course information in the educational information system are transformed into data logs. After a series of data preprocessing operations, we use system clustering to understand the overall clustering result. We apply fast clustering to accurately obtain the clustering results. In this process, a semi-biased statistic is introduced to objectively determine clustering number. By analyzing the learning behavior patterns according to the clustering results, we find that different student groups have different preferences. At the same time, the clustering results are used as the label of the data and the support vector machine is utilized to train data with labels. Three algorithms are used to optimize the parameters in the support vector machine. The constructed model can be applied to predict the future learning behavior. The visual

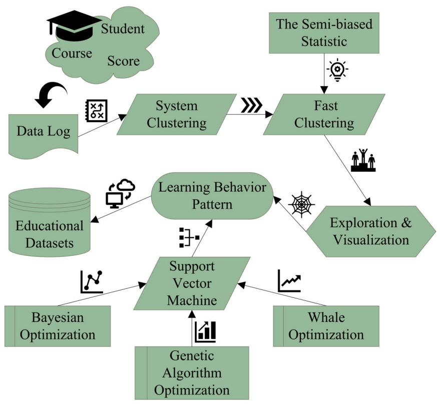


FIGURE 1. **Framework of educational process mining.**

results in every stage are clearly presented to teaching managers for reference in educational management. Therefore, the framework of EPM proposed in this paper is shown in Fig. 1.

III. METHODOLOGY

In dealing with diverse data, the instances in the data can be classified with the help of clustering methods in data mining. We prepare to apply clustering results to the label of training in the classification algorithm [37]. Song *et al.* [38] constructed feature vectors of execution instances from activities. Bose and Van Der Aalst [39] considered the context of the activity, constructed feature vectors based on execution patterns, and then clustered the execution instances using commonly used clustering algorithms. In this paper, students are clustered by calculating the distance among the attributes of multiple courses. On the basis of considering the attributes of the students' courses, the square Euclidean distance is selected as the distance between each course for calculation.

A. SYSTEM CLUSTERING

An overview about the whole clustering result is beneficial to determine number of clusters by using system clustering. Pedigree diagrams that can be drawn by using system clustering present the whole clustering result intuitively. System clustering includes the class average method and Ward's method. Combined with the actual situation, the paper uses Ward's method for research. Firstly, divide n samples into one category, and the sum of squared deviations is 0. Then each time two categories are merged. The sum of squared deviations will increase each time one category is reduced. At this time, two categories with the smallest increase in the sum of squared deviations should be selected for merging until all samples are merged as one category.

For a statistical problem involving in system clustering, there are n samples, and each sample measures p items. Through investigation, the sample observation matrix can be obtained as

$$\mathbf{X} = \mathbf{X}_{ij} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \cdots & \cdots & \cdots & \cdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix} \tag{1}$$

where x_{ij} is the observation data of the jth index of the ith sample. The ith sample is described by the ith row of matrix.

The basic steps of the system clustering method are below.

1) Calculate the squared Euclidean distance between n students to obtain a distance matrix.
2) Initially, n students each form one category, and each category contains only one student. At this time, the intra-class distance is 0, and the inter-class distance is the distance between two students.
3) Merge two categories with the smallest distance to form a new category. At the same time, clustering number is reduced by one.
4) Calculate the distance between the new category and other categories to get a new distance matrix. If the number of merged categories is equal to 1, go to 5), otherwise go back to 3).
5) Draw a pedigree diagram.

B. THE SEMI-BIASED STATISTIC

Applying the proposed semi-biased statistic to determine number of clusters objectively is helpful to analyze the learning behavior patterns from different student groups. In order to better analyze and summarize the learning behavior patterns, the student group is clustered by using the clustering algorithm, and the learning behavior pattern of each category is discussed. Use the fast-clustering K-means algorithm to cluster the students, and use the semi-biased R_k^2 to determine clustering number.

$$\text{The semi-biased } R_k^2 = \frac{B_{KL}^2}{T} = \frac{W_M - (W_K + W_L)}{T} \tag{2}$$

B_{KL}^2 represents the increment of the sum of squares of intra-class deviations after merging categories G_K and G_L as new category G_M. The increment is used to evaluate the effect after combining two categories. T is the sum of the square sum of intra-class deviations and the square sum of inter-class deviations. According to the above definition, the larger the value of the semi-biased statistic in a certain step, the better the clustering effect in the last step.

C. SUPPORT VECTOR MACHINE

Support vector machine (SVM) was proposed by Cortes and Vapnik in 1995 [40], which originated from pattern recognition and is mainly used in classification and regression problems. Assuming that the training set (x_i, y_i), $i = 1, 2, \ldots, l$, $x_i \in R^n$, $y \in \{-1, 1\}$ can be separated by a hyperplane $x\omega+b = 0$. The vector closest to the hyperplane is the largest distance from the hyperplane, then this hyperplane is called the optimal hyperplane. The main idea of SVM is to map the input vector to a high-dimensional feature space through some pre-selected nonlinear mapping, and construct the optimal hyperplane in this space. [41]

SVM adopts the principle of structural risk minimization and is suitable for small sample classification. If it is unknown in advance whether the training set is linearly separable, the objective function and constraint condition can be constructed in the training set.

$$\min \frac{1}{2}\|\omega\|^2 + C\sum_{i=1}^{n} \xi_i \tag{3}$$

$$y_i\left(x_i\omega + b\right) \geq 1 - \xi_i \tag{4}$$

$$\xi_i \geq 0, \quad i = 1, 2, \ldots, l \tag{5}$$

In (3), C is the penalty coefficient, which represents the punishment intensity. ξ_i is a slack variable used to penalize sample points that cannot be accurately classified. C is used to control the balance between sample bias and machine generalization ability. If C is too large, the results of the test set will be overfitted, and if C is too small, the results of the test set will be underfitted. Use Lagrange multiplier method to convert into dual form.

$$\max \sum_{i=1}^{l} \alpha_i - \frac{1}{2}\sum_{i,j} \alpha_i\alpha_j y_i y_j x_i x_j \tag{6}$$

$$\sum_{i=1}^{l} y_i\alpha_i = 0,\ 0 \leq \alpha_i \leq C, \quad i = 1, 2, \ldots, l \tag{7}$$

$$\omega = \sum_{i=1}^{l} y_i\alpha_i x_i \tag{8}$$

If $\alpha_i > 0$, we call the corresponding x_i as support vector. Use the discriminant function $y = \operatorname{sgn}\left[\sum_{x_i \in SV} \alpha_i y_i x_i x + b\right]$ to determine the category of the sample.

SVM is a binary classification model, and the basic model is a linear classifier with the largest interval defined on the feature space. SVM uses kernel tricks, making them essentially non-linear classifiers. We need to discuss which kernel function to choose because different kernel functions correspond to different problems in practical situations. The basic idea of the kernel trick is to map the input space to a feature space through a nonlinear transformation, so that the hyperplane model in the input space corresponds to the hyperplane model in the feature space. The kernel function is shown in (9).

$$K\left(x, z\right) = \varphi\left(x\right)\varphi\left(z\right) \tag{9}$$

$K\left(x, z\right)$ is the kernel function, $\varphi\left(x\right)$ is the mapping function, and $\varphi\left(x\right)\varphi\left(z\right)$ is its inner product. In this paper, the radius basis function (RBF) kernel function with strong anti-interference ability is used as the kernel function. The RBF kernel function can map the input samples in the SVM to the high-dimensional feature space to solve the linear inseparability problem. The kernel function is shown in (10).

$$K\left(x, z\right) = \exp\left(-\frac{\|x - z\|^2}{2\sigma^2}\right) \tag{10}$$

σ controls the radial action range. σ^2 is usually expressed by g. If g is too large, there are too many support vectors. If g is too small, there are fewer support vectors. In order to obtain better classification results, it is necessary to obtain the best parameters C and g through continuous iterations.

D. METRICS FOR EVALUATION

The performance of the classifier is evaluated by using the accuracy, recall, precision and F_1. The calculation methods of the four model evaluation metrics are shown in (11), (12), (13), and (14).

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \tag{11}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{12}$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{13}$$

$$F_1 = \frac{2 \times \text{Recall} \times \text{Precision}}{\text{Recall} + \text{Precision}} \tag{14}$$

Among them, TP represents true positive. The predicted category is positive and the actual category is positive. FP indicates false positive. The predicted category is positive and the actual category is negative. FN denotes false negative. The predicted category is negative and the actual category is positive. TN means true negative. The predicted category is negative and the actual category is negative. For multi-classification problems, the accuracy, recall, precision, and F_1 are calculated by average.

IV. CASE ANALYSIS AND DISCUSSION

Three groups of students who come from different majors are randomly selected in a university, and a dataset is composed of the student data of each group, namely datasets P, Q, and R which have 46, 61, and 51 records respectively. There are 9, 13, and 13 courses respectively. The courses in the same student group and their teachers are the same. Three datasets include student number, courses and grades attributes. The course grade is the direct manifestation of learning behavior. Therefore, this paper analyzes and discusses the learning behavior patterns among three student groups based on curriculum achievement, and summarizes the behavior patterns of the three student groups, so as to provide reference for the general learning behavior patterns. Considering the data we study, we choose SVM which is suitable for small sample classification in order to meet the requirements of the number of records.

For the examples of multiple make-up and re-exams, we choose the principle of using the first valid grade as the attribute value. In this way, it can avoid the high score in the make-up examination affecting the judgment of students'

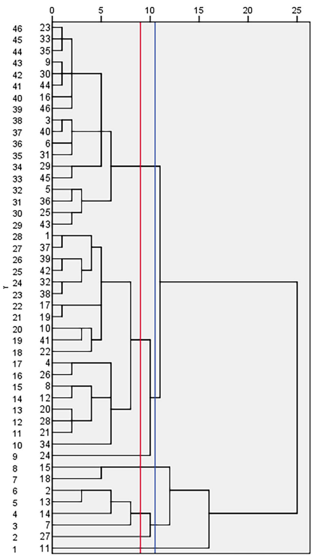


FIGURE 2. The pedigree diagram of dataset *P* using system clustering.

actual learning behavior, and avoid the low estimation of students' learning results caused by the 0-point score due to missing the examination for special reasons. Metadata need to be performed basic preprocessing operations on student grade data to meet the minimum requirements for mining.

A. AN OVERVIEW ABOUT THE WHOLE CLUSTERING RESULT

We can have a general understanding of the overall clustering results of the three datasets through the pedigree diagram that can be drawn by using system clustering. The preprocessed data are clustered according to the connection between groups on the basis of squared Euclidean distance, and the course attribute is used as the input variable. The pedigree diagrams of the three datasets after using system clustering are shown in Figs. 2, 3, and 4 respectively. The first column on the left in Figs. 2, 3, and 4 is counting, and the second column is the student number.

Before discussing the learning behavior patterns by category, the value of the clustering number k ought to be determined. The total number of samples is $n = 46$, $n = 61$, $n = 51$ respectively. According to the empirical rule $k = \lfloor \sqrt{n} \rfloor$, the value of k can take 6, 7, 7. In Fig. 2, the dataset P is divided into 7 clusters according to the red line and 5 clusters according to the blue line. In Fig. 3, the dataset Q is divided into 8 clusters according to the red line and 6 clusters according to the blue line. In Fig. 4, the dataset R is divided into 7 clusters according to the red line. Taking Fig. 4 as an example, there are seven intersections between the red line and the black line. Looking to the left from the intersection, the students included in an intersection belong to the same cluster. When the scribe line moves to the left, the number of intersections increases, that is, clustering number increases. On the contrary, clustering number decreases.

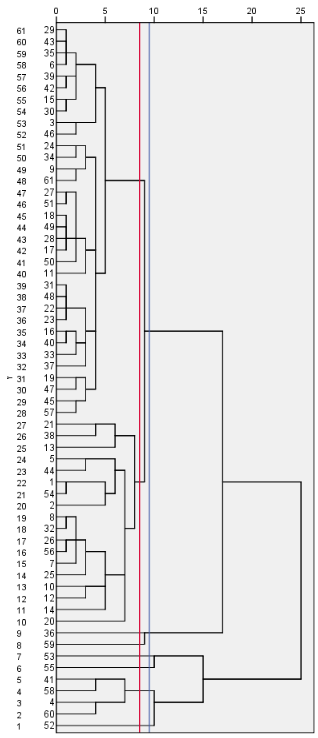


FIGURE 3. The pedigree diagram of dataset *Q* using system clustering.

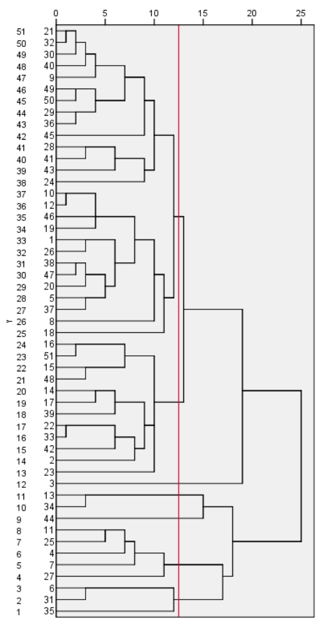


FIGURE 4. The pedigree diagram of dataset *R* using system clustering.

B. USING A SEMI-BIASED STATISTIC TO DETERMINE CLUSTERING NUMBER

In order to further determine clustering number objectively, the preprocessed data are analyzed with the implementation of using the fast-clustering K-means algorithm in which clustering effect corresponding to different clustering numbers can be observed by adjusting the number of clusters. The best k value is between 2 and $\sqrt{n}$ [42], so the maximum k value is $\sqrt{n}$. Starting with two, the intra-class distance and the inter-class distance are obtained at the same time, and the semi-biased statistic can be calculated. The semi-biased R_k^2 evaluates the effect of merging two categories into a new category. The maximum value of k is $\lfloor\sqrt{n}\rfloor - 1$ according to the definition in the semi-biased statistic.

The values of the semi-biased statistic in dataset P are semi-biased $R_{k=2}^2$=0.084867, semi-biased $R_{k=3}^2$=0.105101, semi-biased $R_{k=4}^2$=0.099221, semi-biased $R_{k=5}^2$=0.13579. If k=3, the semi-biased statistic reaches a local maximum value. By comparing semi-biased statistics and combining the pedigree diagram, it is more appropriate to select four as clustering number in dataset P. At this time, the number of iterations is six, and there is no change in the cluster center, so convergence is achieved. Therefore, the first group of students is clustered into four categories.

The values of the semi-biased statistic in dataset Q are semi-biased $R_{k=2}^2$=0.057845, semi-biased $R_{k=3}^2$=0.10064, semi-biased $R_{k=4}^2$=0.096447, semi-biased $R_{k=5}^2$=0.085302, semi-biased $R_{k=6}^2$=0.145626. If k=3, the semi-biased statistic reaches a local maximum value. By comparing semi-biased statistics and combining the pedigree diagram, it is more appropriate to select four as clustering number in dataset Q. At this time, the number of iterations is three, and the cluster center only changes slightly, so convergence is achieved. Therefore, the second group of students is clustered into four categories.

The values of the semi-biased statistic in dataset R are semi-biased $R_{k=2}^2$=0.047031, semi-biased $R_{k=3}^2$=0.068193, semi-biased $R_{k=4}^2$=0.089436, semi-biased $R_{k=5}^2$=0.074164, semi-biased $R_{k=6}^2$=0.11764. If k=4, the semi-biased statistic reaches a local maximum value. By comparing semi-biased statistics and combining the pedigree diagram, it is more appropriate to select five as clustering number in dataset R. At this time, the number of iterations is five, and the cluster center only changes slightly, so convergence is achieved. Therefore, the third group of students is clustered into five categories.

From the calculation results of semi-biased statistics in three datasets, the change trends of the semi-biased statistics are from small to large, then decrease, and finally increase. From the final determination of the clustering number, it is not to take k corresponding to the global maximum of the semi-biased statistic, but to take k corresponding to the local maximum of the semi-biased statistic.

C. USING CROSS VALIDATION TO TEST THE CLUSTERING EFFECT

The clustering result is used as the label of the data to train the data later. Therefore, in order to improve the performance of the subsequent prediction classifiers, it is necessary to test whether the clustering effect is good or not. The quality of the clustering effect will affect the accuracy of analyzing the behavior pattern, so the clustering effect is cross validated.

TABLE 1. Cross validation results of dataset P after clustering using the semi-biased statistic.

	Case	1	2	3	4	Total
Counting	1	14	0	0	0	14
	2	0	16	2	0	18
	3	1	3	6	0	10
	4	0	0	0	4	4
%	1	100.0	0	0	0	100.0
	2	0	88.9	11.1	0	100.0
	3	10.0	30.0	60.0	0	100.0
	4	0	0	0	100.0	100.0

TABLE 2. Cross validation results of dataset Q after clustering using the semi-biased statistic.

	Case	1	2	3	4	Total
Counting	1	25	0	1	0	26
	2	0	26	1	0	27
	3	0	5	2	0	7
	4	0	1	0	0	1
%	1	96.2	0	3.8	0	100.0
	2	0	96.3	3.7	0	100.0
	3	0	71.4	28.6	0	100.0
	4	0	100.0	0	0	100.0

Cross validation was performed on the cases, in which each case was classified by those other than the case, and the validation results are shown in Tables 1, 2, and 3, respectively. It can be found from the three tables that K-means clustering analysis correctly classified 87.0%, 91.8%, and 76.5% of the clustered cases. Without using the semi-biased statistic to cluster the data, the three student groups were clustered into four, five, and four categories, respectively, and the cross-validation accuracy rates were 80.4%, 80.3%, and 86.3%, respectively.

The first student group was clustered into four categories in both cases, but using the semi-biased statistic proposed in the paper to determine clustering number improved the correct rate by 6.6%, and the cases in four categories were different in both cases. It is explained from the side that the semi-biased statistic can not only be used to determine clustering number but also can improve the clustering effect.

The second student group was originally clustered into five categories, but after using the semi-biased statistic to determine clustering number, it was clustered into four categories and the correct rate of cross-validation increased by 11.5%. Objectively determining clustering number made the correct rate greatly improved. Therefore, the results clustered into four categories are selected for analysis.

The third student group was originally clustered into four categories, and the cross-validation accuracy rate was 86.3%. After using the semi-biased statistic, it was clustered into five categories, and the cross-validation accuracy rate was 76.5%. It can be seen from the detailed data in Table 3 that the accuracy of the third and fifth categories is low. In the equal test of the group mean, the significance of most courses of this student group is greater than 0.05, indicating that there is no significant difference in the mean value within

TABLE 3. Cross validation results of dataset *R* after clustering using the semi-biased statistic.

	Case	1	2	3	4	5	Total
Counting	1	5	0	0	1	0	6
	2	0	11	1	1	0	13
	3	0	2	4	1	0	7
	4	0	1	0	11	1	13
	5	0	1	1	2	8	12
%	1	83.3	0	0	16.7	0	100.0
	2	0	84.6	7.7	7.7	0	100.0
	3	0	28.6	57.1	14.3	0	100.0
	4	0	7.7	0	84.6	7.7	100.0
	5	0	8.3	8.3	16.7	66.7	100.0

the category, which explains why the accuracy is reduced after using the semi-biased statistic to determine clustering number. Therefore, the results which are clustered into four categories without using the semi-biased statistic are selected for analysis.

D. EXPLORING THE LEARNING BEHAVIOR PATTERNS IN THREE STUDENT GROUPS

The word pattern is understood from the surface meaning, that is, a form in the development process of things. Different students have different learning behaviors. With the purpose of providing reference for the general learning behavior patterns, we intend to summarize the different learning behavior patterns from different clustering results, that is, the learning behavior of students who belong to the same category can summarize a kind of learning behavior pattern. Table 4 shows the number of various cases, the corresponding student number, and the distances between members in a category and their cluster center in dataset *P*. The last column in the table calculates the average distance between all members in a category and the cluster center. On the premise of large inter-class distance, the goal of clustering is to make the intra-class distance as small as possible. Therefore, the smaller the average distance, the better the clustering effect.

There are 14 students in the first category, which accounts for 30.4% of the total number in dataset *P*, and has the smallest average distance and the best clustering effect. The second category has 18 students, accounting for 39.1% of the total number. There are many students with consecutive numbers in the category. For example, the student numbers are 36, 37, 38, and 39. When arranging dormitories, the university usually arranges them according to student numbers, so students with consecutive student numbers are usually a group of dormitories. The learning atmosphere in the dormitory will affect the students in the whole dormitory, so the students in one dormitory usually cluster into one category. A phrase that appears frequently in recent years, ‘‘Top students’ dormitory’’, verifies the accuracy of the clustering results to a certain extent. There are 10 students in the third category, accounting for 21.8% of the total number. The clustering effect of this category is the worst from the average distance, and the number of students in this category is relatively small. There are 4 students in the fourth category, accounting for 8.7% of the total number. The distance between members in the category and cluster center is quite different, which leads to the poor overall clustering effect of the category.

The number of various cases, the corresponding student number, and the distances between members in a category and their cluster center in dataset *Q* are shown in Table 5. There are 26 students in the first category, accounting for 42.6% of the total number. Except for the fourth category, the category has the smallest average distance and the best clustering effect. The second category has 27 students, accounting for 44.3% of the total number. The consecutive numbers 24, 25, 26, and 27 of the students are all clustered into the second category, which verifies the phenomenon that students in a dormitory are very likely to cluster into the same category. There are 7 students in the third category, accounting for 11.5% of the total number. The cases numbered 52 and 53 in this category have a large distance from the cluster center, resulting in the maximum average distance in dataset *Q*, which leads to poor clustering effect. There is 1 student in the fourth category, accounting for 1.6% of the total number. There is only one student in the category, so this student is the cluster center.

After the dataset *R* uses the semi-biased statistic to determine clustering number, there is no significant difference in the mean in most courses. Therefore, the clustering results obtained by the semi-biased statistic are not used for analysis, but the clustering results with higher accuracy are obtained by cross-validation to discuss the learning behavior pattern in dataset *R*.

The number of various cases, the corresponding student number, and the distances between members in a category and their cluster center in dataset *R* are shown in Table 6. There are 10 students in the first category, accounting for 19.6% of the total number. The category has the smallest average distance and the best clustering effect. The second category has 22 students, accounting for 43.1% of the total number. The consecutive 7, 8, 9, 10, 11, and 12 of the student numbers are clustered into the second category. In the case of 4 to 8 people in a dormitory, the six students must have people in the same dormitory. The overall learning atmosphere in the dormitory affects individuals, so that students in the same dormitory are clustered into the same category. Students in the same dormitory have a great tendency to develop towards the same pattern of learning behavior. There are 11 students in the third category, accounting for 21.6% of the total number. There are many consecutive cases in this category, and the average distance is short. There are 8 students in the fourth category, accounting for 15.7% of the total number. The average distance difference of the four categories in dataset *R* is small, and the clustering effect in the three datasets is the best.

Identifying the given number of clusters means that we can discuss students’ learning behavior by category, and then summarize all kinds of learning behavior patterns. From the perspective of learning patterns, we can identify the student

TABLE 4. Number of various cases, student number and distance between members in a category and their cluster center in dataset *P*.

	Number of cases	Student number	The distance between members in a category and their cluster center	Average distance
The first category	14	3, 6, 9, 16, 23, 29, 30, 31, 33, 35, 40, 44, 45, 46	12.86, 18.18, 8.08, 12.84, 8.76, 14.20, 9.26, 17.01, 11.47, 14.85, 12.14, 8.24, 19.50, 15.84	13.09
The second category	18	1, 4, 5, 10, 17, 19, 20, 24, 25, 28, 32, 36, 37, 38, 39, 41, 42, 43	14.88, 17.42, 18.49, 11.93, 12.68, 18.24, 11.14, 24.14, 20.03, 13.32, 10.05, 12.58, 12.14, 17.73, 20.92, 19.67, 17.04, 18.24	16.15
The third category	10	7, 8, 12, 15, 18, 21, 22, 26, 27, 34	22.62, 18.31, 11.65, 24.07, 21.70, 18.39, 22.52, 18.53, 29.40, 21.91	20.91
The fourth category	4	2, 11, 13, 14	16.10, 22.50, 9.36, 18.59	16.64

TABLE 5. Number of various cases, student number and distance between members in a category and their cluster center in dataset *Q*.

	Number of cases	Student number	The distance between members in a category and their cluster center	Average distance
The first category	26	3, 6, 9, 11, 15, 16, 17, 18, 22, 23, 28, 29, 30, 31, 33, 35, 37, 39, 40, 42, 43, 46, 47, 48, 49, 51	16.85, 13.91, 14.18, 22.52, 14.45, 12.00, 11.31, 12.65, 19.70, 13.07, 12.01, 14.87, 11.26, 15.42, 14.05, 17.01, 17.56, 15.17, 10.45, 15.58, 12.67, 23.18, 20.05, 18.47, 12.60, 16.93	15.30
The second category	27	1, 2, 5, 7, 8, 10, 12, 13, 14, 19, 20, 21, 24, 25, 26, 27, 32, 34, 36, 38, 44, 45, 50, 54, 56, 57, 61	21.54, 23.36, 18.13, 18.04, 19.13, 20.64, 17.74, 26.38, 23.59, 15.66, 24.12, 24.51, 15.15, 23.21, 10.01, 13.97, 15.78, 18.04, 27.71, 19.76, 22.15, 14.33, 19.44, 22.05, 10.29, 13.88, 16.33	19.07
The third category	7	4, 41, 52, 53, 55, 58, 60	17.11, 21.25, 31.69, 29.76, 33.19, 19.90, 20.53	24.78
The fourth category	1	59	0	0

TABLE 6. Number of various cases, student number and distance between members in a category and their cluster center in dataset *R*.

	Number of cases	Student number	The distance between members in a category and their cluster center	Average distance
The first category	10	2, 6, 13, 16, 18, 19, 23, 24, 33, 36	19.58, 20.47, 14.50, 17.80, 10.07, 16.11, 15.19, 14.73, 17.86, 25.60	17.19
The second category	22	1, 3, 4, 5, 7, 8, 9, 10, 11, 12, 14, 15, 17, 21, 22, 25, 26, 27, 28, 31, 32, 34	20.96, 16.39, 15.97, 17.14, 21.84, 19.49, 18.06, 11.16, 13.45, 22.41, 12.74, 16.95, 23.34, 22.05, 13.66, 16.87, 17.49, 22.69, 13.46, 18.25, 19.90, 14.15	17.66
The third category	11	20, 29, 30, 35, 41, 44, 45, 47, 48, 49, 51	14.55, 19.49, 14.50, 14.83, 14.49, 21.52, 26.14, 17.74, 10.02, 13.84, 22.39	17.23
The fourth category	8	37, 38, 39, 40, 42, 43, 46, 50	27.53, 18.75, 22.14, 11.84, 19.00, 21.73, 21.04, 22.28	20.54

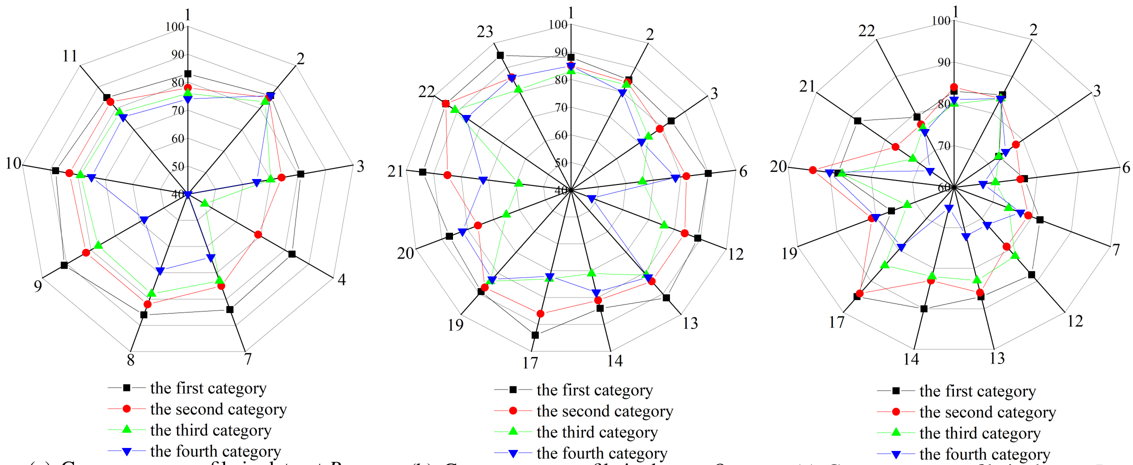
groups according to the actual situation of students, which can be divided into several categories for the analysis of learning patterns. From the perspective of educational process, academic performance reflects the learning process. By studying academic performance, we can understand the learning process and provide guidance for improving students' behavior in the learning process.

In the real settings, teachers will rank the students according to the absolute score of the course, but criteria for ranking are subjective to some degree and do not match the students' actual mastery of the course. We identify number of clusters based on the students' actual situation about their course grade. Therefore, these groups of students correspond to the real settings.

E. USING VISUALIZATION TECHNOLOGY TO PRESENT LEARNING BEHAVIOR PATTERNS

The final cluster center of clustering analysis can be used to describe the characteristics of learning behavior patterns in various categories. As a visual representation form, the presentation of profile needs to rely on data visualization technology. Geometric graphics commonly used in the field of education such as histogram, radar chart and pie chart can better support the presentation of text-based data and profile. The student profiles can be constructed according to the clustering results. Students' course grades are the most direct reflection of course mastery. Radar charts are mostly used for comprehensive presentation of multi-index data, so radar charts are utilized to construct student profiles. Since students are clustered according to their course grades, the final student profiles are the group course profiles of students.

Three groups of students are randomly selected in a university, and a dataset is composed of the student data of each group. A dataset is clustered into several categories. A radar chart reflects the courses' mastery of different categories of students in a dataset, which means we can make a comparison among different categories in a dataset. A radar chart also demonstrates overall courses' mastery of this student group, which represents overall courses' mastery from different datasets can be compared. The course profiles of



(a) Group course profile in dataset P (b) Group course profile in dataset Q (c) Group course profile in dataset R

FIGURE 5. Course profiles of three datasets.

TABLE 7. Accuracy, recall, precision, F_1 and corresponding optimal parameters of using three optimization methods and unoptimized results.

		Train_accuracy	Test_accuracy	Recall	Precision	F_1	C	g
Dataset P	NO	37.50%	42.86%	42.86%	18.37%	25.71%		
	BO	100%	92.86%	92.86%	93.88%	92.62%	27.25	0.00885
	GAO	100%	85.71%	85.71%	85.71%	85.71%	19.42	0.01409
	WOA	96.88%	92.86%	92.86%	93.88%	92.62%	48.81	0.00398
Dataset Q	NO	47.62%	31.58%	31.58%	9.97%	15.16%		
	BO	95.24%	94.74%	94.74%	96.49%	95.14%	19.02	0.00153
	GAO	100%	94.74%	94.74%	95.18%	94.61%	20.46	0.09575
	WOA	90.48%	100%	100%	100%	100%	5.49	0.00375
Dataset R	NO	37.14%	56.25%	56.25%	31.64%	40.50%		
	BO	100%	93.75%	93.75%	96.88%	94.17%	24.93	0.03774
	GAO	100%	93.75%	93.75%	96.88%	94.17%	19.42	0.01409
	WOA	97.14%	100%	100%	100%	100%	7.28	0.00651

three datasets are shown in Fig. 5. The category closer to the periphery in the radar chart, the better the overall learning situation, and vice versa. It can be found from the figure that the overall learning situation of the first category in dataset P is the best, the second category is better than the third category, and the fourth category has the worst overall learning situation. In dataset Q, the overall learning situation of the first category is the best, the second category is better than the third category, and the fourth category has the worst overall learning situation. In dataset R, the overall learning situation of the first category is the best, the second category is better than the third category, and the fourth category has the worst overall learning. In Fig. 5, compared with (a) and (b), the overall fluctuation range of (c) is small, ranging from 60 to 100. Combined with the calculation of the average distance above, it can be concluded that the better the overall learning situation of the students, the better the clustering effect of the category which they belong to.

The construction basis of the profile is consistent with the concept of higher education development, and has strong practical guiding significance. First of all, universities and teachers can inquire about student information based on the profiles, so that teachers can clearly understand the course mastery of various students and make timely changes to the teaching plan, so as to achieve the purpose of teaching students according to their aptitude. Secondly, after the student profiles are determined, books can be recommended for different categories of students according to the students' learning behavior habits. Finally, although student profiles may change over time, teachers can predict students' future academic performance based on the mastery of the current course, paving the way for improving the overall level of students.

F. PERFORMANCE OF SUPPORT VECTOR MACHINE CLASSIFIER

The performance of the classifier constructed by SVM is validated by random hold-out method. The random hold-out method divides the original data into three categories: training set, validation set, and test set. The training set is used to learn and train model, and the validation set is utilized to adjust the parameters to determine the best model. The test set is applied to test and evaluate the performance of the classifier. In this paper, Bayesian Optimization (BO), Genetic Algorithm Optimization (GAO) and Whale Optimization Algorithm (WOA) are used to determine the optimal parameters respectively,

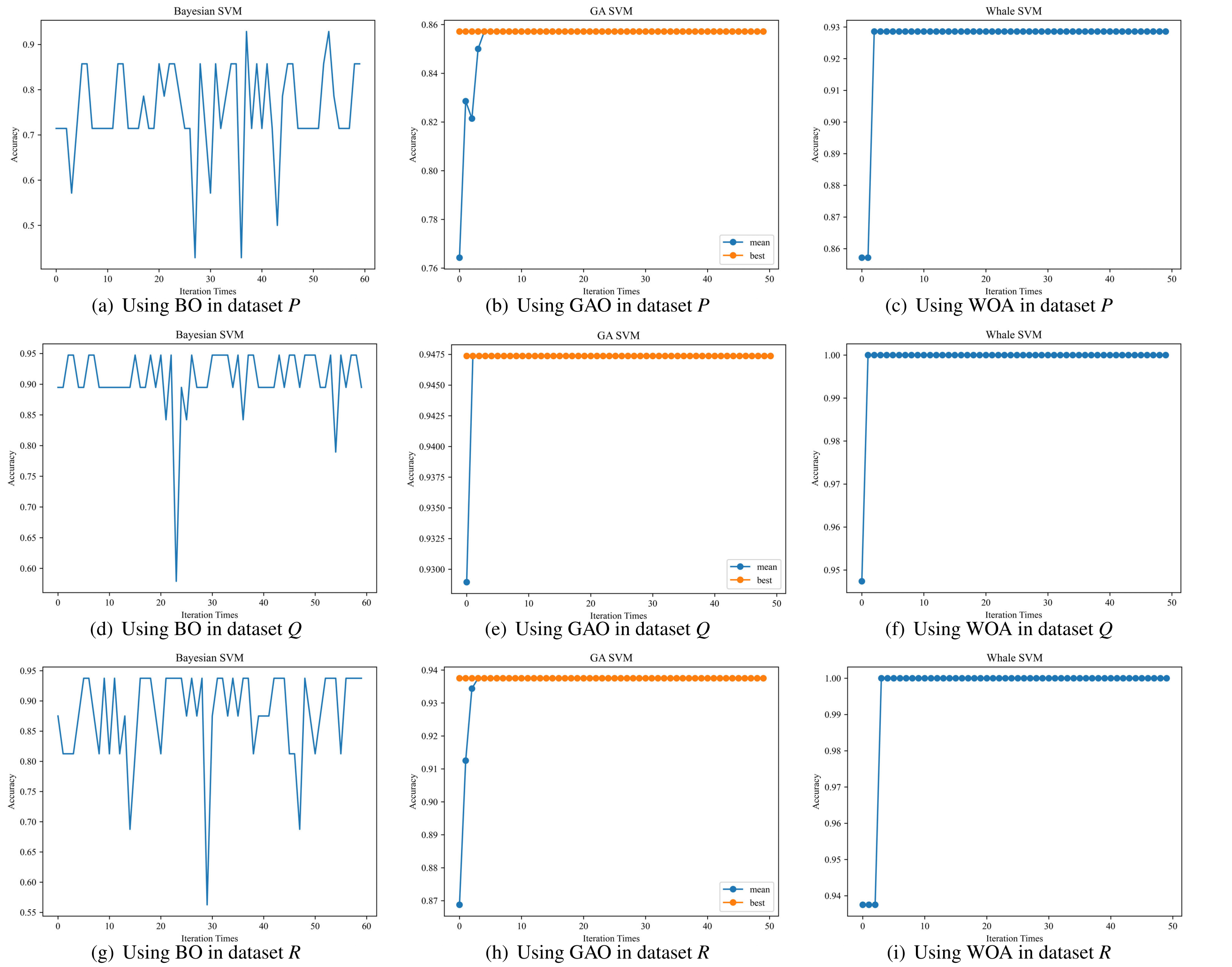


FIGURE 6. The trend of the accuracy of the test set with the number of iterations is shown in the figure. The first, second, and third columns are the results of BO, GAO, and WOA, respectively. The first, second, and third row are the variation trends of the accuracy rates in datasets *P*, *Q*, and *R* after three optimizations, respectively.

so the original data is directly divided into training set and the test set. The training set accounts for 70% and the test set accounts for 30%. Since the random state cannot be determined, there is an optimal set of parameters for each run in BO. The optimal parameters for GAO and WOA are unique. The unoptimized and three optimized results and their corresponding optimal parameters are shown in Table 7. NO in the table stands for unoptimized, and the unoptimized parameters are randomly determined. It can be found from the table that compared with the unoptimized results of randomly determined parameters, the optimized accuracy, recall, precision and F_1 have been significantly improved. Among the three methods for optimizing parameters, WOA works best.

After using three methods of optimizing parameters respectively, the trend of the accuracy of the test set in three datasets with the number of iterations is shown in Fig. 6. In Fig. 6, (a), (b), and (c) are the results of the dataset P after three optimizations. Dataset P uses WOA with fewer iterations and higher accuracy. In Fig. 6, (d), (e), and (f) are the results of the dataset Q after three optimizations. Dataset Q uses GAO and WOA with fewer iterations, but WOA has higher accuracy. In Fig. 6, (g), (h), and (i) are the results of the dataset R after three optimizations. Dataset R uses GAO and WOA with fewer iterations, but WOA has higher accuracy.

In Fig. 6, (a), (d), and (g) are the results of BO. Combining the data in Table 7, it can be found that the accuracy of the test set is the peak value. In Fig. 6, (a) achieves the optimal prediction effect after nearly 40 iterations. In Fig. 6, (d) and (g) achieve optimal predictions multiple times during the process. Compared with the other two optimizations, the results of BO are less stable. The main reason is that the uncertainty of the random state leads to the non-unique optimal parameters. In Fig. 6, (b), (e), and (h) are the results of GAO. The blue line represents the historical forecast mean. The orange line is the best predicted result. Three datasets achieve optimal prediction results within 10 times. In

Fig. 6, (c), (f), and (i) are the results of WOA. Three datasets achieve the optimal prediction effect within 5 times. Compared with the other two optimizations, WOA has less iterations, more stable trends and higher accuracy.

V. CONCLUSION AND FUTURE WORK

From system clustering to constructing predictive classifiers, from proposing a semi-biased statistic to quantitatively determine clustering number to using three algorithms to optimize parameters in SVM, from exploring learning behavior patterns by category to constructing student courses group profiles, the EPM framework proposed in this paper provides decision support for educators and shows learning behavior habits to the educated. EPM makes the black box of education process display information and knowledge that are beneficial to the development of universities and students, and has strong practical guiding significance for both educators and educated people. We comprehensively apply DM and PM to study the data in the educational process to find some interesting discoveries in order to give some suggestions to stakeholders. For example, students can know their academic level and improve their grades, teachers can make teaching plans according to the actual situation of students, and teaching managers can improve training programs to adapt to students' abilities.

As a technology emerging in the business field, PM shows great potential in the application of education. For example, test scores, document writing and course mining are all the applications of process mining technology in the field of education. EPM pays more attention data mining in the educational process. Therefore, the emergence of EPM has expanded the boundaries of EDM. Students' family background, book borrowing, daily consumption and other information can be used as data sources. The data used in this paper is the data exported from the educational information systems, and the time-related characteristics are not taken into account. If the time-related characteristics can be associated with learning behavior, it would be a new scalable direction. Compared with big data, this paper uses SVM which is suitable for small sample classification, so the technology used in EPM is not limited to system clustering and SVM used in this paper. Since the development of PM, corresponding software and special algorithms [3], [43], [44] have been developed. Researchers can learn from them to study EPM and make contributions for education.

REFERENCES

[1] W. M. P. Van Der Aalst, T. Weijters, and L. Maruster, "Workflow mining: Discovering process models from event logs," *IEEE Trans. Knowl. Data Eng.*, vol. 16, no. 9, pp. 1128–1142, Sep. 2004.

[2] C. D. S. Garcia, A. Meincheim, E. R. Faria Junior, M. R. Dallagassa, D. M. V. Sato, D. R. Carvalho, E. A. P. Santos, and E. E. Scalabrin, "Process mining techniques and applications—A systematic mapping study," *Exp. Syst. Appl.*, vol. 133, pp. 260–295, Nov. 2019.

[3] M. R. Dallagassa, C. Dos Santos Garcia, E. E. Scalabrin, S. O. Ioshii, and D. R. Carvalho, "Opportunities and challenges for applying process mining in healthcare: A systematic mapping study," *J. Ambient Intell. Humanized Comput.*, vol. 13, no. 1, pp. 165–182, Jan. 2022.

[4] N. Trcka and M. Pechenizkiy, "From local patterns to global models: Towards domain driven educational process mining," in *Proc. 9th Int. Conf. Intell. Syst. Des. Appl.*, 2009, pp. 1114–1119.

[5] H. Awatef, B. Gueni, M. Fhima, and S. David, "Process mining in the education domain," *Int. J. Adv. Intell. Syst.*, vol. 8, nos. 1–2, pp. 1–14, 2015.

[6] J. E. Cook and A. L. Wolf, "Automating process discovery through event-data analysis," in *Proc. 17th Int. Conf. Softw. Eng. (ICSE)*, Seattle, WA, USA, 1995, pp. 73–82.

[7] R. Agrawal, D. Gunopilos, and F. Leymann, "Mining process models from workflow logs," in *Proc. 6th Int. Conf. Extending Database Technol.*, Berlin, Germany, 1998, pp. 469–483.

[8] W. M. P. Van Der Aalst, *Process Mining: Data Science in Action*. Berlin, Germany: Springer-Verlag, 2016.

[9] J. De Weerdt, A. Schupp, A. Vanderloock, and B. Baesens, "Process mining for the multi-faceted analysis of business processes—A case study in a financial services organization," *Comput. Ind.*, vol. 64, no. 1, pp. 57–67, Jan. 2013.

[10] M. Werner, "Financial process mining—Accounting data structure dependent control flow inference," *Int. J. Accounting Inf. Syst.*, vol. 25, pp. 57–80, May 2017.

[11] M. Borkowski, W. Fdhila, M. Nardelli, S. Rinderle-Ma, and S. Schulte, "Event-based failure prediction in distributed business processes," *Inf. Syst.*, vol. 81, pp. 220–235, Mar. 2019.

[12] M. Zayoud, Y. Kotb, and S. Ionescu, "β Algorithm: A new probabilistic process learning approach for big data in healthcare," *IEEE Access*, vol. 7, pp. 78842–78869, 2019.

[13] N. Martin, J. De Weerdt, C. Fernández-Llatas, A. Gal, R. Gatta, G. Ibáñez, O. Johnson, F. Mannhardt, L. Marco-Ruiz, S. Mertens, J. Munoz-Gama, F. Seoane, J. Vanthienen, M. T. Wynn, D. B. Boilève, J. Bergs, M. Joosten-Melis, S. Schretlen, and B. Van Acker, "Recommendations for enhancing the usability and understandability of process mining in healthcare," *Artif. Intell. Med.*, vol. 109, Sep. 2020, Art. no. 101962.

[14] N. Martin, L. Pufahl, and F. Mannhardt, "Detection of batch activities from event logs," *Inf. Syst.*, vol. 95, Jan. 2021, Art. no. 101642.

[15] E. De Roock and N. Martin, "Process mining in healthcare—An updated perspective on the state of the art," *J. Biomed. Informat.*, vol. 127, Mar. 2022, Art. no. 103995.

[16] C. Diamantini, L. Genga, F. Marozzo, D. Potena, and P. Trunfio, "Discovering mobility patterns of Instagram users through process mining techniques," in *Proc. IEEE Int. Conf. Inf. Reuse Integr. (IRI)*, San Diego, CA, USA, Aug. 2017, pp. 485–492.

[17] D. Myers, S. Suriadi, K. Radke, and E. Foo, "Anomaly detection for industrial control systems using process mining," *Comput. Secur.*, vol. 78, pp. 103–125, Sep. 2018.

[18] Y. Huang, C. Zhao, G. Zhao, J. Liu, F. Shu, and H. Li, "Intelligent technology of education process mining: Research framework, status and prospect," *e-Educ. Res.*, vol. 41, no. 8, pp. 49–57, 2020, doi: 10.13811/j.cnki.eer.2020.08.007.

[19] S. Anuwatvisit, A. Tungkasthan, and W. Premchaiswadi, "Bottleneck mining and Petri net simulation in education situations," in *Proc. 10th Int. Conf. ICT Knowl. Eng.*, Bangkok, Thailand, Nov. 2012, pp. 244–251.

[20] N. Trcka, M. Pechenizkiy, and W. M. P. Van Der Aalst, "Process mining from educational data," in *Handbook of Educational Data Mining*. London, U.K.: CRC Press, 2010, pp. 123–142.

[21] V. Southavilay, K. Yacef, and R. A. Calvo, "Process mining to support students' collaborative writing," in *Proc. 3rd Int. Conf. Educ. Data Mining*, Pittsburgh, PA, USA, 2010, pp. 257–266.

[22] A. Bogarín, R. Cerezo, and C. Romero, "A survey on educational process mining," *WIREs Data Mining Knowl. Discov.*, vol. 8, no. 1, p. e1230, Jan./Feb. 2018.

[23] M. Bannert, P. Reimann, and C. Sonnenberg, "Process mining techniques for analysing patterns and strategies in students' self-regulated learning," *Metacognition Learn.*, vol. 9, no. 2, pp. 161–185, Aug. 2014.

[24] H. Ariouat, A. H. Cairns, K. Barkaoui, J. Akoka, and N. Khelifa, "A two-step clustering approach for improving educational process model discovery," in *Proc. IEEE 25th Int. Conf. Enabling Technol. Infrastruct. Collaborative Enterprises (WETICE)*, Paris, France, Jun. 2016, pp. 38–43.

[25] A. R. C. Maita, M. Fantinato, S. M. Peres, L. H. Thom, and P. C. K. Hung, "Mining unstructured processes: An exploratory study on a distance learning domain," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Anchorage, AK, USA, May 2017, pp. 3240–3247.

[26] J. Maldonado-Mahauad, M. Pérez-Sanagustín, R. F. Kizilcec, N. Morales, and J. Munoz-Gama, ''Mining theory-based patterns from big data: Identifying self-regulated learning strategies in massive open online courses,'' *Comput. Hum. Behav.*, vol. 80, pp. 179–196, Mar. 2018.
[27] L. Juhaňák, J. Zounek, and L. Rohlíková, ''Using process mining to analyze students' quiz-taking behavior patterns in a learning management system,'' *Comput. Hum. Behav.*, vol. 92, pp. 496–506, Jan. 2019.
[28] S. S. Modi and S. B. Jagtap, ''Multimodal web content mining to filter non-learning sites using NLP,'' in *Proc. Int. Conf. Comput. Netw. Big Data IoT (ICCBI)*, Madurai, India, 2018, pp. 23–30.
[29] J. P. Salazar-Fernandez, J. Munoz-Gama, J. Maldonado-Mahauad, D. Bustamante, and M. Sepúlveda, ''Backpack process model (BPPM): A process mining approach for curricular analytics,'' *Appl. Sci.*, vol. 11, no. 9, p. 4265, May 2021.
[30] K. Akhuseyinoglu and P. Brusilovsky, ''Exploring behavioral patterns for data-driven modeling of learners' individual differences,'' *Frontiers Artif. Intell.*, vol. 5, Feb. 2022, Art. no. 807320.
[31] A. Cooper, *The Inmates are Running the Asylum: Why High-Tech Products Drive us Crazy and How to Restore the Sanity*. Indianapolis, IN, USA: Sams, 2004.
[32] E. Trandafili, A. Allkoci, E. Kajo, and A. Xhuvani, ''Discovery and evaluation of student's profiles with machine learning,'' in *Proc. 5th Balkan Conf. Informat. (BCI)*, Novi Sad, Serbia, 2012, pp. 174–179.
[33] K. Park, H. Ji, and H. Lim, ''Development of a learner profiling system using multidimensional characteristics analysis,'' *Math. Problems Eng.*, vol. 2015, Jan. 2015, Art. no. 652623.
[34] P. Dwivedi and K. K. Bharadwaj, ''E-learning recommender system for a group of learners based on the unified learner profile approach,'' *Exp. Syst.*, vol. 32, no. 2, pp. 264–276, Apr. 2015.
[35] H. Li, X. Zhang, X. Li, Y. Wang, Y. Ling, and D. Mu, ''Construct the students' consumption portrait based on large-scale campus data,'' in *Proc. 12th Int. Conf. Mach. Learn. Comput.*, Shenzhen, China, Feb. 2020, pp. 198–201.
[36] X. Li and S. He, ''Research and analysis of Student portrait based on campus big data,'' in *Proc. IEEE 6th Int. Conf. Big Data Anal. (ICBDA)*, Xiamen, China, Mar. 2021, pp. 23–27.
[37] G. Feng, M. Fan, and Y. Chen, ''Analysis and prediction of students' academic performance based on educational data mining,'' *IEEE Access*, vol. 10, pp. 19558–19571, 2022.
[38] M. Song, C. W. Gunther, and W. M. P. Van Der Aalst, ''Trace clustering in process mining,'' in *Proc. Bus. Process. Manage. Workshops*. New York, NY, USA: Springer-Verlag, 2009, pp. 109–120.
[39] R. P. J. C. Bose and W. M. P. Van Der Aalst, ''Trace clustering based on conserved patterns: Towards achieving better process models,'' in *Proc. Bus. Process. Manage. Workshops*. New York, NY, USA: Springer-Verlag, 2010, pp. 170–181.
[40] C. Cortes and V. N. Vapnik, ''Support vector networks,'' *Mach. Learn.*, vol. 20, no. 3, pp. 273–297, 1995.
[41] V. N. Vapnik, *The Nature of Statistical Learning Theory*. New York, NY, USA: Springer-Verlag, 1995.
[42] M. R. Rezaee, B. P. F. Lelieveldt, and J. H. C. Reiber, ''A new cluster validity index for the fuzzy c-mean,'' *Pattern Recognit. Lett.*, vol. 19, nos. 3–4, pp. 237–246, Mar. 1998.
[43] C. W. Gunther and W. M. P. Van Der Aalst, ''Fuzzy mining-adaptive process simplification based on multi-perspective metrics,'' in *Proc. Int. Conf. Bus. Process. Manage.* Berlin, Germany: Springer, 2007, pp. 328–343.
[44] C. Li, M. Reichert, and A. Wombacher, ''The minadept clustering approach for discovering reference process models out of process variants,'' *Cooperat. Inf. Syst.*, vol. 19, nos. 3–4, pp. 159–203, Sep./Dec. 2010.

GUIYUN FENG was born in Nantong, China, in 1996. She received the bachelor's degree in information management and information system from Nantong University, Nantong, in 2019. She is currently pursuing the master's degree with the School of Management, Guizhou University, Guiyang, China. Her research interests include educational process mining, data mining, pattern recognition, and service science and innovation management.

MUWEI FAN (Member, IEEE) was born in Guiyang, China, in 1987. He received the B.S. degree in electronic engineering from the Beijing Institute of Technology, China, in 2009, the M.S. degree in computer integrated manufacturing from Nanyang Technological University, Singapore, in 2012, and the Ph.D. degree in petroleum storage and transportation engineering from the China University of Petroleum, China, in 2017. Since 2017, he has been an Associate Professor with the Center of Big Data Creative Innovation and Technology, Guizhou University. His research interests include complex system modeling and simulation, reliability analysis and efficiency evaluation, big data analysis, and pattern recognition.

CHUCHU AO was born in Tianmen, China, in 1990. She received the bachelor's degree in international economics and trade from Wuhan Textile University, Wuhan, in 2014. She is currently pursuing the M.B.A. degree with Guizhou University. Her research interests include performance evaluation of organization, behavioral analysis, and operations research.