

Research on Personalized Learning Model Based on Learning Objectives in The Context of Educational Big Data

Qunfeng Wei* and Bin Qi

Zhejiang Business Technology Institute, Ningbo 315012, China

Abstract

After three years of COVID-19, Internet-based online education has developed very well, such as online Moocs, flipped classroom, etc. Such large-scale online courses provide good support for teaching resource sharing and information exchange. However, for the increasingly large online education data, its effective research and utilization is not deep enough, especially the in-depth discussion of personalized learning in the big data environment is not sufficient enough. In order to optimize the quality of online education, strengthen the tracking intervention and supervision of online learning, and provide guiding instructional suggestions, this paper proposes an analytic model of online learning behavior. Using the unique large amount of learning data generated through learning, the online learning behavior is studied and analyzed, and the results are applied to the online learning platform. We design the online learning behavior analysis model in the big data environment, and construct the learning behavior analysis model from left to right and top down. In longitudinal data acquisition, use method, process analysis and application, in turn in the horizontal from based on K-Means algorithm learning resources using cluster analysis, based on A priori learning behavior and learning performance of association rules analysis, based on knowledge error rate group volume strategy, based on TAN network learning style dynamic prediction of four aspects, and based on cluster analysis and association rules give enlightenment, according to the group volume strategy and style prediction effect.

Keywords

big data; Online learning behavior analysis; Learning style prediction; Data mining; TAN Bayesian network.

1. Introduction

1.1. Research background and significance

1.1.1. Research background

In recent years, the development of big data mining technology and mature to traditional education injected fresh blood, online learning is developing rapidly, learners of learning resources personalized and practical requirement is more and more high, people began to use big data mining and analysis technology from the huge amounts of data information found use value of information, provide suitable learning resources push service, so as to better meet the demand of online learning, realize humanization and personalized teaching service. Now big data mining technology has been widely and successfully applied to quantitative investment, financial risk control, accurate marketing, health care, crime prediction and biology, and achieved great success, but in the application of remote education has not made significant achievements, in the current learning system, such as learning source information, students use resource information, student platform search records, students comment information, the course answer information and test practice information, etc., but not effective use of these information. Research big data scholars believe that these data can provide personalized

learning services for learners, through data learning, analysis and mining can find hidden learning behavior information, provide different levels of users, its application mainly monitoring and analysis, prediction and intervention, evaluation and feedback, intelligent guidance and adaptive, etc., use of big data learning data, application learning analysis, course and teaching to make reasonable decisions[1]. With the combination of big data and distance education, the realization of real personalized learning is becoming imminent.

1.1.2. Purpose of research

The ultimate goal of teaching is to promote learners' learning [2], And the personalized learning environment is created to achieve this purpose. The realization of personalized learning by experts and learners from huge amounts of education data can reflect the student learning change law of useful data, thus learning analysis, on the basis of scholars' learning interest, learning style, learning behavior data accurately provide learners personalized learning path, provide learners with targeted personalized learning resources, practice test, and learning advice and feedback, customized for their own development of personalized online courses, achieve effective learning purposes [3].

From the aspects of education personnel, need to pass teaching feedback to analyze the students 'learners characteristics elements, found that conventional and unconventional learning mode, analyze the difficulty of the course blindspot, choose efficient education teaching activities, reasonable correction curriculum structure and content, at the same time educators can be inspired from the students' feedback, broaden the teaching ideas, improve the disadvantages existing in the traditional teaching.

1.1.3. Significance of studying

Personalized learning conforms to the education concept of their aptitude, also can reflect the information age to the learner centered teaching idea, realize the principle of fair and open, reflect the every academic personality development, broke the learners in the limitation of learning time and space, meet the demand of diversified learning, embodies the essence of education.

In terms of theory, it is helpful to study the nature of learning and the psychology of learners; In terms of practical application, it is helpful to track learners 'learning process, evaluate learners' learning effect, and accurately grasp the learning state, In order to enable an early intervention, To improve the learning efficiency and learning quality; From the perspective of educational resources, it improves the utilization rate of resources, Reduce useless junk data and misleading learning; For the students, Improve the resource utilization and also improve the quality of learning, And let the learners to learn to have a strong interest in learning; For the teachers, It can improve the teaching efficiency and is more conducive to the analysis of the disadvantages existing in the teaching; For teaching, research and course developers, To facilitate the evaluation of curriculum structure and learning efficiency, Through various aspects of the feedback, To provide a clear guidance direction for curriculum reform and innovation; For the teaching institutions, Application of enhanced data mining in teaching, Can improve the efficiency of customized courses, Make the future development path of teaching more clear.

1.2. Research status

Online learning in foreign countries, it is called E-learning. Online learning platform provides learners with a large number of online learning resources, breaks the limitation of learning time and space, and meets the diversified learning needs of learners. In recent years, the typical representative of online learning platform is the large-scale online open course —— MOOC [4].

From the development of foreign online education learning platform, in 2000, the UK invested 50 million pounds to fund online education in the UK; in 2001, MIT plans to upload all

courseware materials to provide free use online [5]; In 2002, UNESCO further proposes to open educational resources (Open Educational Resources, OER), Through communication technology, provide educators and learners with various educational and learning resources based on free access, reference and application; In 2004, DFES confirmed that personalized learning consists of five basic elements: learning evaluation, effective teaching, curriculum selection, school management, curriculum expansion and support; In 2008, the George Siemens of the University of Manitoba and the National Research Council of Canada jointly launched a course on "Associated Attention and Associated Knowledge" [6], They believed that an open course that publishes both participants and learning resources on the Internet would be more effective for learning [7]; In 2011, Stanford University professor Sebastian Thrun founded Udacity, his first free online learning course called Introduction to Artificial Intelligence. 160,000 students enrolled, and about 15 percent of online students completed Stanford level courses [8]; In 2012, Harvard University and MIT jointly created the edX platform to provide quality online courses to learners in an open and free format[8]; Similar online education learning platforms also include Khan Academy (Khan Academy), Canvas, Udacity, Coursera, and Venture Lab, etc. This new form of convenient, open and participatory online education has attracted human research interest [9]. Online learning platform recorded the learning interaction between teachers and learners of behavior, produced a large number of continuous teaching-learning interactive information, the information is the learners scattered, unconscious learning behavior, is the depth of learning behavior and learning style, learning dynamic favorable material, through the analysis can reflect the most real learning thinking and situation.

From the perspective of literature research, in 2003, Hummel K and other scholars used a network learning system based on web page technology to collect and analyze their learners' login log and background database, and then studied the characteristics of the learning behavior of network learners [10]; In 2007, Graf S et al. proposed that the field learning style scale could realize the recommendation of personalized learning resources for users [11]; In 2014, William Ribarsky visual vision was the main way of big data analysis. They studied the competitive advantages of data mining in social media analysis by integrating interactive visualization and automated analysis methods and the resulting applications [12]. To sum up, it can be seen that the research direction of foreign online learning focuses more on the extraction of online learners' characteristics and related learning resources, which provides a guiding light for our research direction.

With the rapid development of domestic education informatization in recent years, more and more attention has been paid to personalized learning in the big data environment, and some achievements have been made in theory and research, but there are still shortcomings. At present domestic research on personalized learning mainly has two ways: one is through the data analysis in the process of online learning, the other is to use the learners before test and learning of dynamic data analysis of a personalized feature extraction, but these two ways are not fully considering the learners original cognitive structure and learning style, make the accuracy of personalized characteristics is not enough [13] In addition, the data acquisition of learning behavior lacks pertinence, so how to obtain valuable learning data is an important basis for personalized learning mining. Therefore, this paper proposes to classify the learning style of online learners; cluster the learning behavior by cluster analysis, find out the implicit learning rules in the learning behavior of learners, and recommend personalized papers and personalized courses. Through the study of personalized learning model under the big data environment, the shortcomings in the research under educational information in China are gradually made up.

1.3. Study content and research methods

1.3.1. Research content of the paper

First, the research background, purpose and significance of the paper are described, and the research method and organization of the paper are clarified. Then we do the theoretical analysis. Starting from the theoretical basis of personalized learning under the big data environment, the design of personalized learning under the environment of big data model, the design using K-Means cluster analysis technology to analyze the use of online learning resources, put forward improved Apriori association rule analysis method, using genetic algorithm to generate personalized practice, etc., and design how to according to the online learners learning behavior predicts the learning style of learners. And summarizes the characteristics and improvements of personalized learning model in the big data environment, and makes prospects for further research on related topics.

1.3.2. Research technique

(1) Literature review method. The research literature on education big data, online learning platform, learning style, learning behavior, personalized papers and personalized courses were collected, sorted out, summarized and analyzed, clarify the research direction and research content, and make basic preparation for the research of this paper.

(3) Participate in the observation method. In the process of analyzing the online learning platform, by participating in the observation of the platform learning, the functions of watching videos, learning tests and participating in discussions in the online learning platform are analyzed, so as to understand the specific significance of the collected data.

(4) Content research method. According to the existing needs of educational big data mining theory, learning style, learning behavior science theory and practice, the personalized learning model design under the big data environment is proposed, and the learning research platform is used to realize the model and observe its changes.

2. Construction of An Online Learning Behavior Analysis Model in A Big Data Environment

On the basis of the analysis of online learning behavior, on the basis of learner characteristics theory, learning analysis theory and learning style theory, build the online learning behavior data structure framework and data collection, the collected data from top to down, from left to right learning behavior analysis and modeling, from online learning cluster analysis, online learning association analysis, online learning style analysis three parts model building horizontal analysis, in the longitudinal data selection and processing, analysis process and result output design, and finally the application of the results. As shown in Figure 1, the overall architecture of online learning behavior analysis model design in the big data environment is presented.

2.1. Online learning and group volume association strategy

In the syllabus will require the assessment of the knowledge points to grasp the degree, generally have the following levels: mastery, proficiency, mastery, understanding. In this paper, these levels are quantified as 1,2,3, and 4 weights. In this paper, the tree structure is used to represent the knowledge point graph, and the knowledge point management of the tree structure is shown in Figure 2.

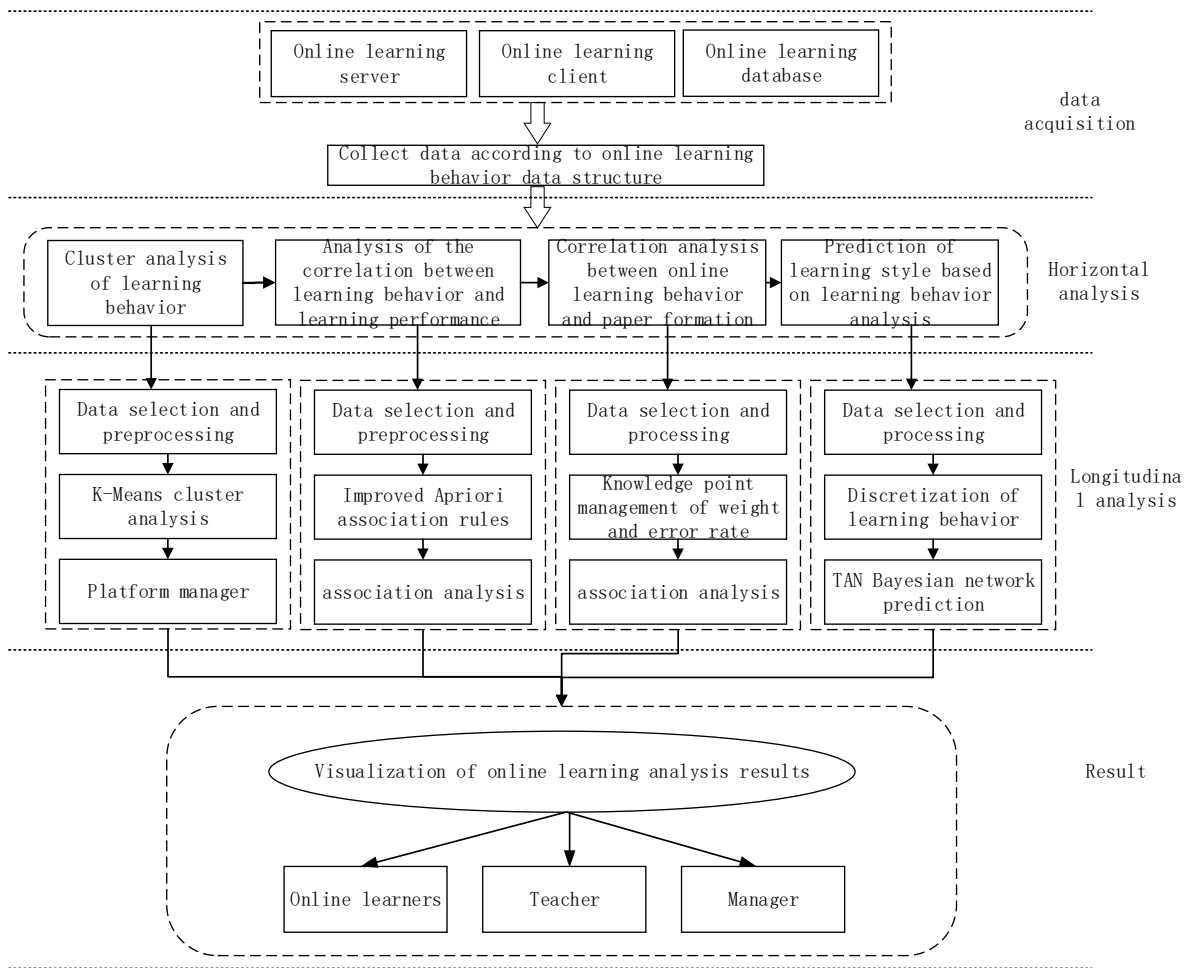


Figure 1. Flow chart of personalized learning model design

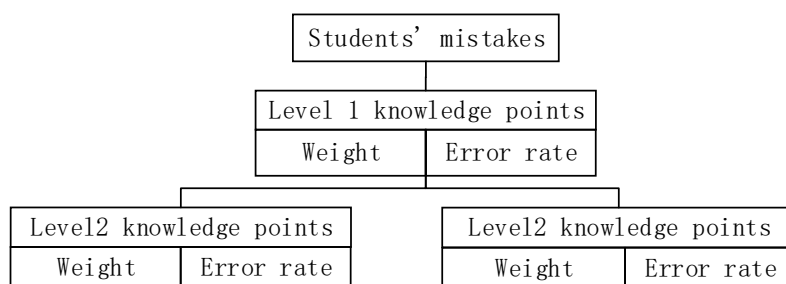


Figure 2. Tree knowledge point management structure based on the association of knowledge point weight and error rate

The error rate element is added to the tree knowledge point management structure, and the weight of the new knowledge points is calculated. Select the proportion to increase the proportion, so as to improve the extraction rate, as shown in formula (1):

$$W'_i = W_i + (1 + y_i) * p_i \tag{1}$$

$W_i y_i p_i$ This is the weight of the i knowledge point, the error rate of the student in the i knowledge point, and the weight proportion of the i knowledge point in his brother node. In the selected questions, the knowledge points with high error rate is to improve the priority of these knowledge points, focus on these knowledge points with high error rate, can automatically select the blind spot easy to error point.

The constraints of the selection are: the total score of exercises, the total time of answering, the constraints of knowledge points, the distribution of questions and the distribution of the difficulty of the paper. According to these five constraints, a $P=N \times 7$ target matrix is established, where N is the number of questions for a set of exercises, as shown in the following formula: $P = N \times 7$

$$P = \begin{bmatrix} a_{11} & \cdots & a_{1m} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nm} \end{bmatrix} \quad (m=7) \quad (2)$$

Each line in this matrix represents a topic selection, and each topic has 6 attributes that are represented by the target matrix column: the ID number of the question in the database = a_{i1} , the question type = a_{i2} , the question score =, the associated knowledge point =, the question difficulty level =, the best answer time of the question = a_{i6} , the chapter of the question belongs =. $a_{i3} a_{i4} a_{i5} a_{i7}$

Through the collection and sorting of wrong questions, learners provide personalized simulation practice papers for their own learning blind spots and weak points. The specific steps are shown as follows:

Step 1: Build the tree structure of the knowledge points in each chapter, as shown in the figure, according to the mastery of the knowledge points

Distribution weight;

Step 2: Supplement the knowledge points of the tree structure according to the error rate of the learner in the problem;

Step 3: Calculate the new weight of the knowledge points through formula (1), and arrange the knowledge points in descending order according to the new weight;

Step 4: Suppose there is m channel in each chapter and n first-level knowledge points, set $a=m/n$, $b=m \bmod n$, select the first $a+1$ second-level knowledge points in the first b first-level knowledge points of this chapter, and select the first a second-level knowledge points in the second $n-b$ first-level knowledge points;

Step 5: If the section is yet to be selected, perform step 3;

Step 6: Determine the practice structure, including the total score, the number of each question type, the score constraint of each question type, and other cross the questions, and finally select the questions that meet the conditions.

2.2. Diagnosis of the learning style based on learning behavior

In this paper, Felder-Silverman learning style model is adopted, while ILS scale is a pre-measurement table with strong subjective consciousness and the real results may be biased. Therefore, 36 teare selected to generate the training dataset of pre-test data, and then calculate validity and predict learning style by comparing the training dataset with learning behavior.

2.2.1. Static diagnosis design of the learning style based on the learning scale

According to the previous section of this paper Felder-Silverman learning style model each learners divided into four dimensions eight types, various types of style behavior characteristics also made specific description, researchers from these classification, for example, visual learners to provide video information, more tasks about multimedia type to

them, for active learners teachers organization group cooperation to enable them to actively discuss, and appropriate to think learners provide more thinking space and time, etc.

In order to initialize the learning style, Felder and Soloman according to the FLSM learning style model designed Solomon learning style scale ILS, recognized by many scholars in the industry, and successfully applied with some famous adaptive learning system, proved that the scale in the network teaching environment explicit determine learning style has good applicability and validity [14]. Therefore, this paper will use the ILS style scale to preset the learning style and apply it to the learning style training data set.

ILS style scale has 44 questions, each question has a and b two options, starting from each dimension corresponding to the two kinds of learning style, set and style of learning behavior related problems, these topics contains the four dimensions of learning style classification, and the topic of each dimension as shown in table 1 cross arrangement, each dimension corresponds to 11 questions, according to the status of the students answer to judge the students learning style type.

Table 1. The Solomon Learning Style Scale

Active / contemplative type			Inentional / intuitive			Visual type / speech type			The sequence / comprehensive type		
question	a	b	question	a	b	question	a	b	question	a	b
1	1		2		1	3	1		4		1
5		1	6	1		7	1		8		1
9		1	10	1		11	1		12		1
13		1	14		1	15	1		16	1	
17	1		18		1	19		1	20	1	
21		1	22		1	23	1		24	1	
25		1	26	1		27	1		28		1
29	1		30	1		31	1		32	1	
33		1	34	1		35		1	36	1	
37		1	38		1	39	1		40	1	
41	1		42		1	43		1	44	1	
amount to	4	7	amount to	5	6	amount to	8	3	amount to	7	4
(Large-smaller) + larger letters)											
3b			b			5a			3a		

The analysis of the ILS Learning Style Scale has the following rules:

1. The style orientation of the information processing dimension is judged by the selection results of questions 1,5,9,... 41, and the learning style of the perceptual dimension is judged by questions 2,6,10,... 42, etc., among which the letters represent the type of learning style;
2. Fill in the real situation according to the questions on the questionnaire and complete the above form, select the question "A" and fill in "1" in the "A" column and select "b" and fill in "1" in the "b" column;
3. Calculate the sum of each column and fill it in the corresponding total column;
4. In each style dimension, subtract the smallest total number with the largest total value, and write down the letters of the difference value and the partial large value, and fill in both in the last column in the form of "Xa". The larger the value before the letter, the stronger the style.

It is worth noting that, The final value for each dimension can only be the " 11a, 9a, 7a, 5a, 3a, a,b, 3b, 5b, 7b, 9b, One value in the 11b ", The larger the coefficient before the letter, It shows that the style characteristics are too strong, The smaller the coefficient indicates that the

weaker personality traits, Will mix with some characteristics of the opposite character, Therefore, this paper will further divide these 12 values into three parts based on the empirical research and literature reference, "11a, 9a, 7a, 5a ", " 3a, a,b, 3b ", " 5b, 7b, 9b, 11b", The middle part is balanced, With partial characteristics on both sides, Take the first dimension, for example, The learning style characteristics of this dimension {active type, balance type, reflective style}.€

2.2.2. Recommendation based on the learning style

According to the characteristics of the learning style, different learning paths, learning methods and learning materials are recommended to the learners, as shown in Table 2 below.

Table 2. Learning recommendation table based on learning style

Style dimension	type	Feature performance	Recommended results
Information processing dimension	Active type	Like to discuss, try to do things	Forum (required), reading learning materials (recommended), watching learning videos (recommended), doing exercises (required), completing tests (required)
	reflective style	I like to think quietly, study and work independently	Read learning materials (required), browse the outline and knowledge tree (required), forum discussion (recommendation), watch learning videos (required), do exercises (required), complete tests (required)
Information perception dimension	Feeling type	Like specific things	The reading attributes are concrete learning materials (required), watching learning videos (required), forum discussion (recommended)
	intuition type	Like the abstract things	Read attributes for abstract learning materials (required), forum discussion (recommended)
Information input dimension	visual type	Like the media type resources	Browse the knowledge tree (required), watch learning videos (required), browse pictures (required), forum discussion (recommended)
	verbal type	Like the textual resources	Browse the course syllabus (required), browse the knowledge tree (recommended), read the text type resource materials (required), forum discussion (required)
Content understanding dimension	Serial type	Like the linear learning knowledge	Knowledge tree (required), course video (required), unit exercise (required)
	synthesizing type	Like comprehensive learning knowledge	Curriculum outline (required), forum discussion (required), comprehensive exercise (required)

2.2.3. T A N The Bayesian classification dynamically determines the learning style

Using the Bayesian network to speculate that the learning style is an effective calculation method. When the Bayesian network is used to detect students' learning style, the nodes include the learning style dimension and the behavior mode of the dimensions, in which the parent node is the behavior mode of each dimension of the learning style, and the child node is each dimension of the learning style. According to the FLSN learning style model, the relevant behavior patterns of each dimension can be extracted, so as to establish the TAN Bayesian Network (T ree Augmented Naive Bayesian). Because TAN has the dependence between variables, the learning algorithm is needed to construct the model structure.

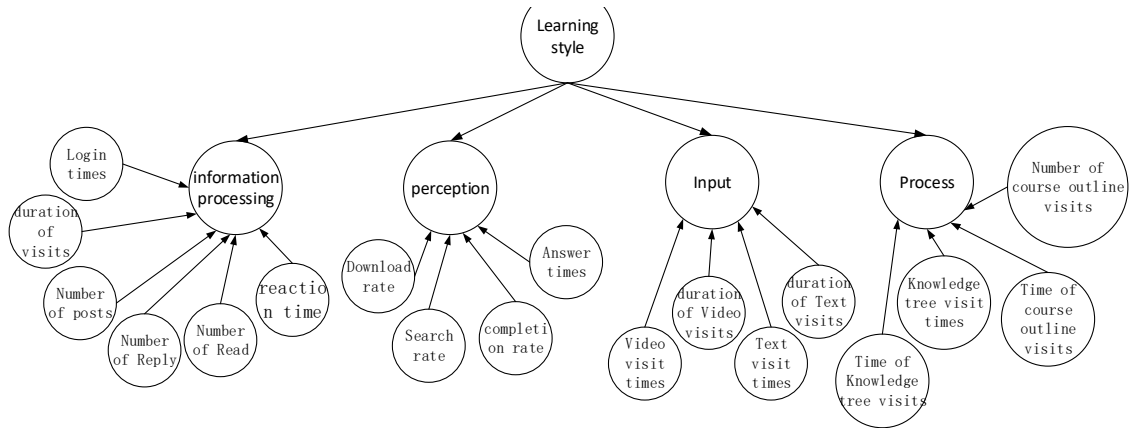


Figure 3. Learning style network based on TAN

(1) Calculate the mutual information value between the attribute variables. The mutual information value between the attribute variables is calculated by the attribute characteristics in the sample set. The mutual information value between all the pairs of the learning behavior attribute variables is calculated separately in the four dimensions. Each learning behavior attribute value is set to, that is, the learning behavior attribute X is defined under the known dimension CX_i , $X_i \in \{L, M, H\}$; And X_j Conditional mutual information between the two states.

$$I_C(X_i; X_j|C) = \sum_{X_i, X_j, C} P(X_i; X_j; C) \log_2 \frac{p(X_i; X_j|C)}{p(X_i|C)p(X_j|C)}, i \neq j \tag{3}$$

(2) Generate a weighted and completely undirected graph. According to the mutual information value in step 1, the value is filled in the completely undirected graph as shown in FIG. 4 to establish the weighted completely undirected graph for the $arI_C(X_i; X_j|C)$

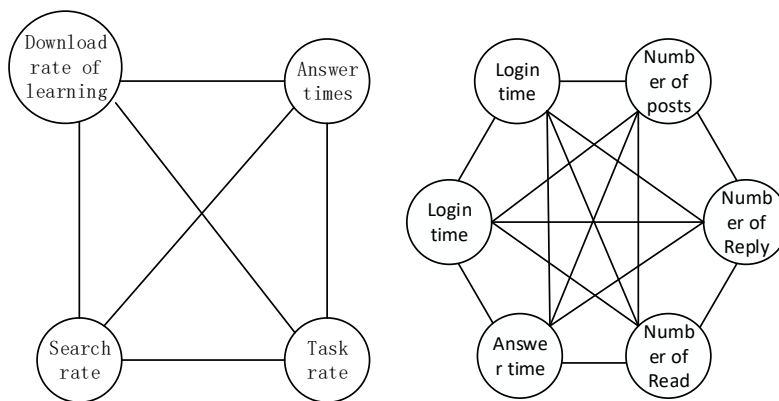


Figure 4. Learning style part attributes completely undirected graph

(3) Construct the maximum weight tree. In different dimensions, the mutual information values of each attribute pair are sorted in descending order, and the nondirectional edges are deleted in turn to construct the maximum weight tree following the principle of no loop until $n-1$ edges are selected. Then any attribute node is selected as the root, and the outward direction of the root node is the direction between the attribute nodes.

(4) Add the total classification attribute parent node to complete the Bayesian network structure construction.

(5) After constructing the Bayesian network, the conditional prior probability of each node is generated according to the network structure. Firstly, because the root node is independent, there is no other parent node except the parent node, so the prior probability of the root node is, and the conditional prior probability of other nodes is $P_{C_k}(X_i)P_{C_k}(X_i|X_k)$

C_k (6) Predictive learning style. Using the learning style $k=1,2$, $k=1$ indicates the left learning style, $k=2$ represents the right learning style, by calculating the probability formula of the learning style:

$$P(X|C_k) = \prod_{i=1}^n p(X_i|C_k) \quad (4)$$

Where n is the number of learning behaviors that affect the learning style.

The learner's learning style is preliminarily determined by the following formula (5), and the learning style orientation is finally determined by the value size calculated by formula (3):

$$P(C = C_i) = \frac{P(X|C_i)}{P(X|C_i)+P(A|C_{3-i})} 100\% \quad (5)$$

Due to the strong and weak learning styles, the ILS described in literature [34] due to FLSM divides the various dimensions of the detected learning styles into several grades 1,3,5,7,9,11 [34] 50% – 66.7%66.8% – 83.4%83.5% – 100% Among them, 1,3 represent balance; 5,7 indicate weak learning style; 9,11 indicate learning style with strong upward trend and extreme style. Therefore, this paper divides the probability of 50% -50% 100% into three levels corresponding to the ILS learning style level, corresponding to the balanced type, indicating that the upward trend of learning style is weak, and the learning style is strong. Balanced learners have the characteristics of both sides of the learning style, and they belong to the most stable class of students. In the process of online teaching, students should be guided to be balanced as far as possible.

2.3. Conclusion

The learning behavior analysis model in the big data environment analyzes the learning behavior from the perspectives of horizontal correlation analysis, cluster analysis and prediction, and models the learning behavior analysis from the longitudinal data collection, method selection and application process.

3. Testing and Implementation of Personalized Learning System in The Big Data Environment

3.1. Dataset preparation

This paper uses the background data of the online course platform of a university online research institute as the main source of analysis. The platform has been running smoothly for several years, and has accumulated 19,914 online student records and tens of thousands of other traffic data. At present, it obtains knowledge through online learning, so it has accumulated very large data resources.

Among the data sheet is the learner basic characteristics data sheet PE_User, course information data sheet Course, resource data information sheet PE_Resource, student homework test information sheet PE_UserTest, and forum Thread, etc. According to the characteristics of the students constructed above, this paper, the data sheet is divided into several categories as shown in Figure 5.

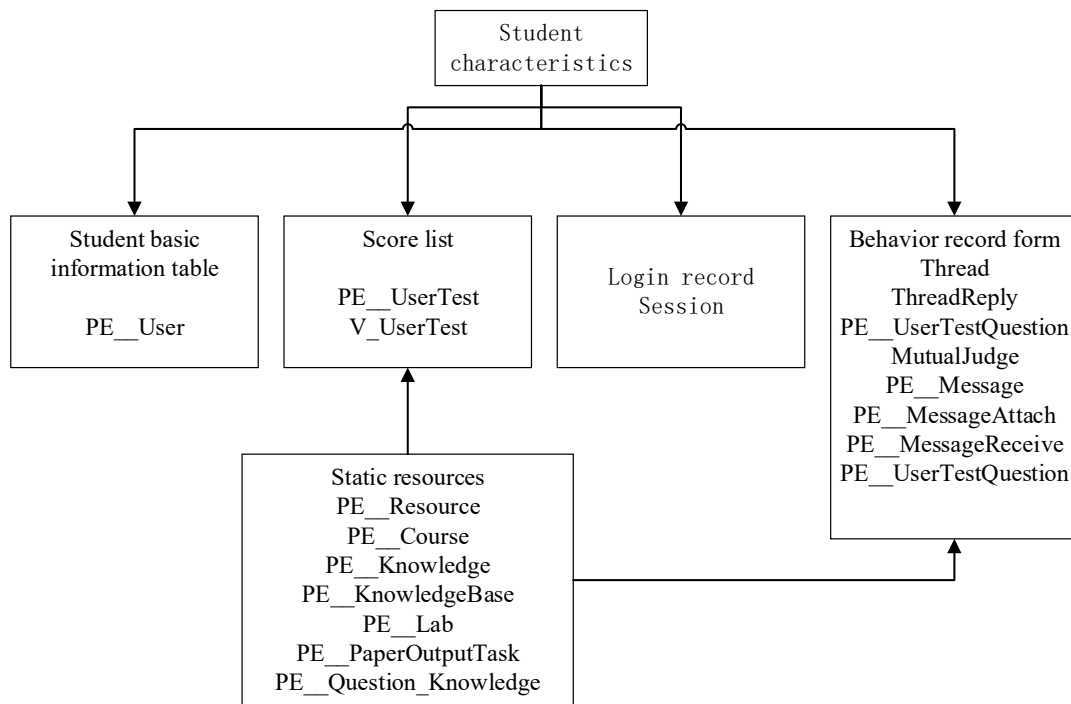


Figure 5. Data table corresponding to students' characteristics

According to the above learning behavior data model of the online learners, according to the learners different module learning elements of the data, due to different types of online learning behavior data, such as structured type, unstructured type and semi-structured types, at the same time, the collected data sources may be redundant, incomplete, noisy, noise to data before data analysis, and heavy, missing and conversion, the specific processing process is shown in Figure 6.

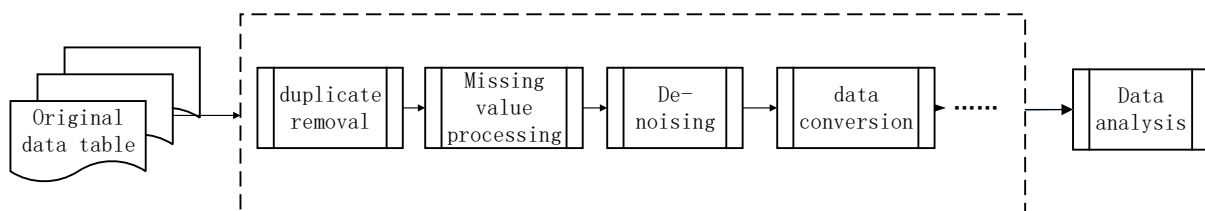


Figure 6. Data pre-processing process of the online learning platform

3.2. Online learning by behavioral cluster analysis

This paper for stored in SQL Server database online learning platform of 40127 traffic data preprocessing, user online learning resources are diverse, according to the reality of the platform to select the most frequent seven resources, online reading, data, BBS, video resources, experimental exercises, homework and practice, set K-means cluster analysis K value of 7, as shown in figure 7.

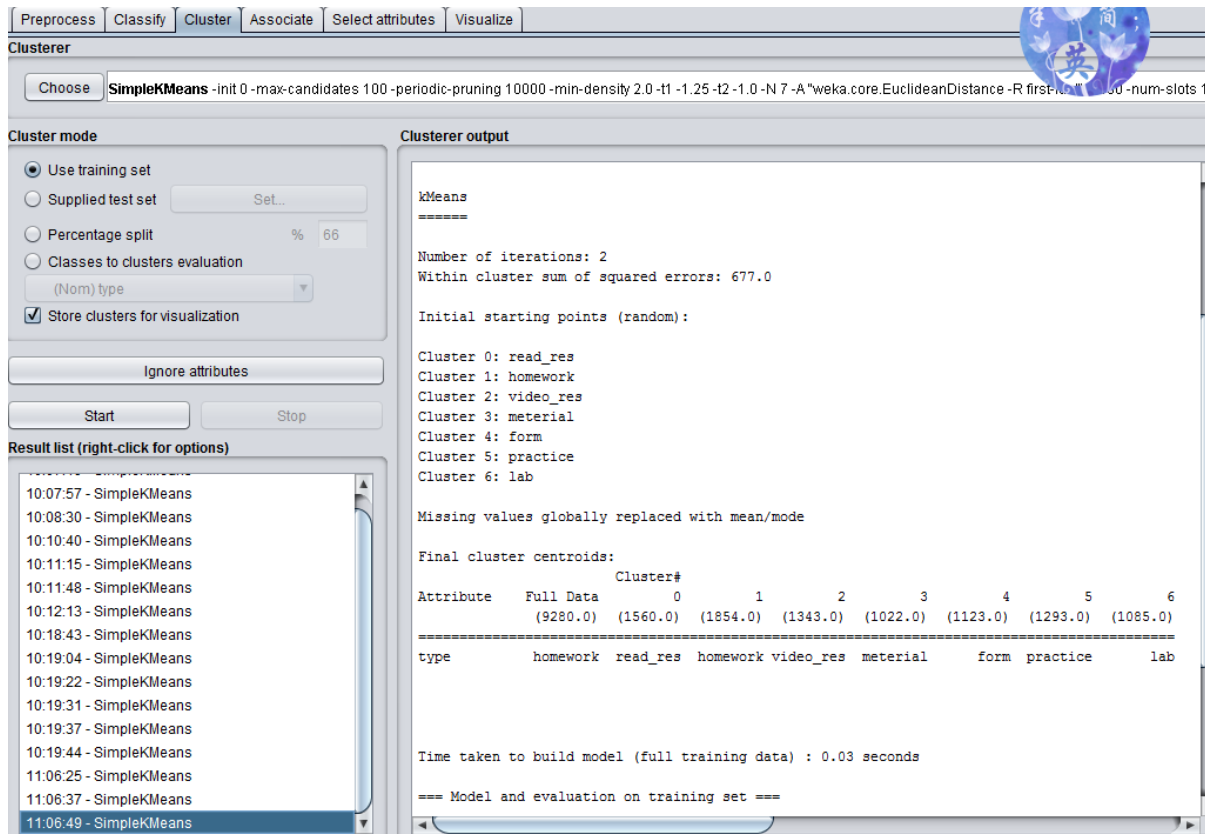


Figure 7. K-Means cluster analysis of the learning resource results

According to the analysis results, a three-dimensional pie map of the distribution of online resource learning as shown in Figure 8 is drawn. It can be seen that the most important online learning methods in this platform are video, as well as homework exercises and experiments, followed by tasks for forum discussion, data download, online reading, etc. Based on the proportion of online courses studied, Online teaching resources should be set up as follows: (1) increase the share of video courses, Research shows that the online course is not only saving time and cost, It is also helpful to the students to review; (2) Online learners spend 20% of their experiences on homework completion, Although there is online intelligent marking, However, the teacher should still solve the problem in time; (3) Learners have also invested a lot of energy in the forum discussions, To give full play to the communication work of the forum, On the one hand, it is conducive to the communication of colleges, On the other hand, through the forum discussion and q & A can solve the problems that the learners encounter in the learning process, Teachers should often communicate with learners online; (3) The learning materials of online learning platforms may not be sufficient, Learning materials are an important bridge for learners' learning progress; (4) Learners also spend more time on online reading, Reading online, in addition to the text, And the knowledge tree, The Syllabus, etc., Knowledge tree helps learners comb through their knowledge, And the syllabus helps to grasp the mastery of the course, Online teaching platforms should provide more knowledge about this aspect.

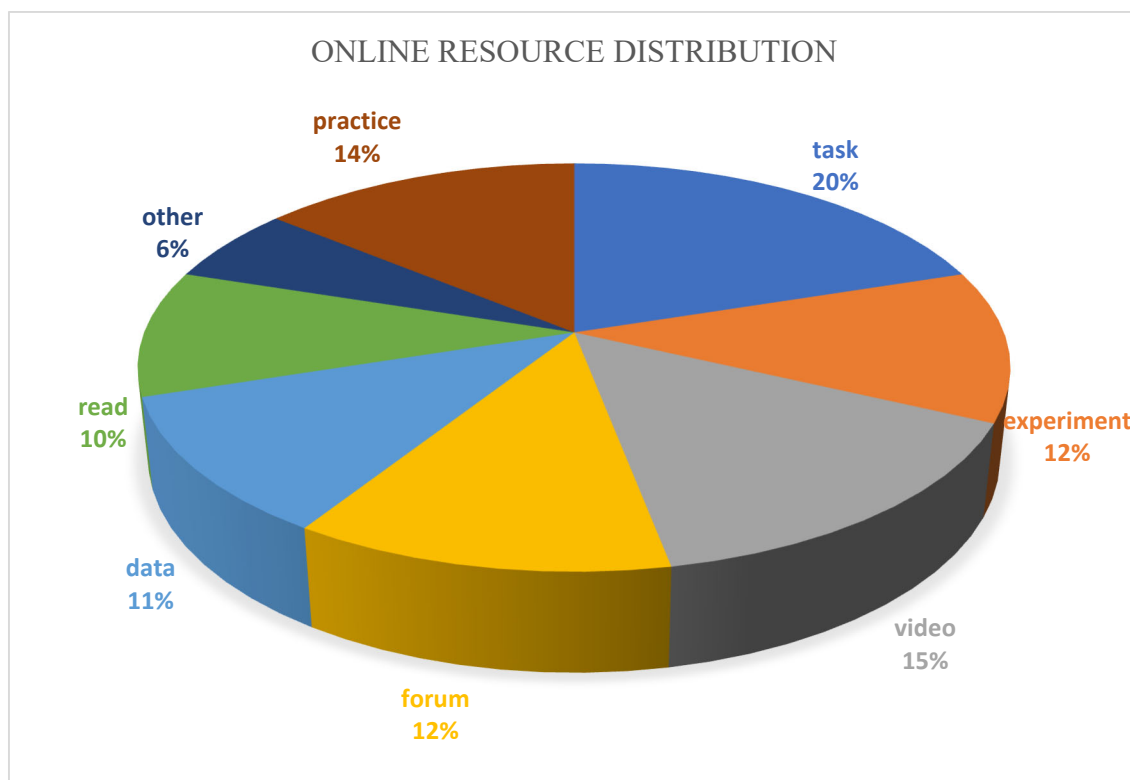


Figure 8. 3 D pie chart of online resource learning distribution

3.3. Learning style detection based on learning behavior

3.3.1. Preset learning style

In order to avoid the burden and inconvenience caused to students by the large number of problems in the learning scale, This paper adopts the preset learning style in the initial learning stage, Literature [15] summarizes the analysis of network learning style: learners are more reflective in the dimension of information processing type, More intuitive bias in the perceptual dimension, More biased visual type in the input dimension, In the dimension of understanding, While the literature [16] shows that, through a large number of experiments, Due to the particularity of their majors for computer science majors, Their learning style is positive in the information processing dimension, The rest are all the same, Therefore, this paper will preset the computer professional learners in the initial stage as {active, intuition type, visual type, Sequential type}, However, the learning style of the computer major is preset to {meditation, intuition type, visual type, Sequential type}.

3.3.2. Learning-style detection based on the TAN Bayesian network

The learning behaviors of learners in the online learning platform of this paper mainly includes: logging in the platform, learning courses, information searching, browsing the web, accessing learning resources, watching course videos, forum discussion, completing homework, completing tests, exams, mutual evaluation, etc.

According to the threshold in Section 3.2.2, each feature attribute of learning style is divided into {L, M, H}, which represents the intensity of learning style characteristics. As shown in Table 3, the learning behavior data training set of the information processing dimension selected in the learning platform. proj 1 represents the learning style category on the left, and proj 2 represents the learning style category on the right.

Table 3. Learning style training dataset

ID	A 1	A 2	A 3	A4	A 5	A6	Proj
1	H	M	H	H	M	H	proj 1
2	L	L	M	L	L	H	proj 2
3	L	L	M	L	L	M	proj 2
4	M	H	M	L	H	M	proj 2
5	L	M	M	M	L	M	proj 2
6	L	L	H	L	L	M	proj 2
7	L	L	L	M	L	M	proj 2
8	M	M	M	L	L	M	proj 2
9	M	M	M	L	L	M	proj 2
10	H	H	H	M	L	L	proj 1
11	M	H	H	L	L	M	proj 1
12	L	L	M	L	L	H	proj 1
.....
37	H	M	H	H	M	M	proj 1
38	H	H	M	L	L	L	proj 1

$P(A/B) = P(AB)/P(B)$ A prior probability table of the information processing dimension is calculated according to the conditional probability formula. As shown in the following table:

Table 4. The CTP table for the Proj

P roj	price
proj1	12/38
proj2	26/38

Table 5. The prior probability table for the attribute nodes

Information processing dimension																		
attribut e	Number of login systems			Visit the forum for longer			Posting amount			Back post quantity			Read the post quantity			Answer time		
scope	L	M	H	L	M	H	L	M	H	L	M	H	L	M	H	L	M	H
proj1	0	2/12	10/12	0	4/12	8/12	0	2/12	10/12	2/12	3/12	7/12	5/12	5/12	2/12	3/12	8/12	1/12
proj2	11/26	10/26	5/26	10/26	12/26	4/26	12/26	12/26	2/26	21/26	5/26	0	23/26	1/26	2/26	3/26	22/26	1/26

The prior rates of learning style for the other dimensions are shown in Table 6:

3.4. Brief summary

Using Section 2 applying the learning behavior analysis model to practice, select the online learning platform independently developed by the institute to analyze and summarize the use of learning resources; analyze the experiment of learning behavior and the correlation of learning performance and obtain the teaching enlightenment; fault in the learning process, generate the personalized practice papers meeting the personal situation; and predict the learning style through learning behavior analysis based on TAN Bayesian network.

Table 6. Prior probabilities of each attribute node in other dimensions

Information perception dimension												
attribute	Download rate of learning materials			Search rate			Homework completion rate			Number of answers		
scope	L	M	H	L	M	H	L	M	H	L	M	H
sens1	2/18	6/18	10/18	4/18	11/18	3/18	2/18	1/18	15/18	2/18	5/18	11/18
sens2	7/20	12/20	1/20	7/20	10/20	3/20	5/20	10/20	5/20	3/20	16/20	1/20
Information input dimension												
attribute	Video access duration			The number of video visits			Text resource access duration			Number of text resource visits		
scope	L	M	H	L	M	H	L	M	H	L	M	H
input Sty 1	0	10/15	5/15	1/15	7/15	7/15	9/15	6/15	0	2/15	11/15	2/15
inputSty2	7/13	6/13	0	5/13	6/13	2/13	1/13	9/13	3/13	1/13	7/13	5/13
Content understanding dimension												
attribute	Knowledge tree browsing time			Number of knowledge tree visits			The duration of the course syllabus browsing			Number of course syllabus visits		
scope	L	M	H	L	M	H	L	M	H	L	M	H
co m preSty 1	2/15	8/15	5/15	1/15	8/15	6/15	7/15	7/15	1/15	9/15	4/15	2/15
compreSty2	16/23	7/23	0	13/23	9/23	1/23	12/23	6/23	5、23	9/23	10/23	4/23

4. Summary and Outlook

4.1. Full text summary

Big data environment, the implementation of the personalized learning model for platform managers, on the one hand, can according to the actual situation of analysis and research, to adjust the platform, so as to improve the efficiency of the platform, on the other hand, according to the analysis of online learning behavior, can help platform managers real-time monitoring platform user loyalty and design a more humanized personalized learning platform.

Innovative research in this paper is as follows:

- (1) Predict the learning style based on the TAN Bayesian network based on the learning behavior, and put forward the personalized recommendation of the learning style according to the learning style.
- (2) According to the correlation between the topic and the knowledge points, we propose a personalized simulation practice implementation model based on the weight of the knowledge points and the error rate.
- (3) The improved Apriori calculation is used to realize the correlation rule extraction between learning behavior and learning performance.
- (4) Through the specific implementation method of online learning behavior analysis, the clustering analysis and correlation analysis of online learning behavior in the big data environment, which provides relevant suggestions and enlightenment for learners' learning evaluation in the platform and teachers to improve the learning method.

4.2. Future expectations

In this paper, we deeply studied the analysis of online learning behavior, summarized and analyzed the characteristics of learning behavior and visually display the analysis results, and constructed a learning style diagnostic model based on such analysis results. According to the shortcomings in the research, the follow-up work and prospects are as follows:

- (1) As the platform may not be mature and perfect in the developer, and has limited data types in behavioral data collection, such as no online real-time dialogue. Therefore, in the future,

more attention will be paid to big data collection to improve the universality of the application of personalized learning model

(2) this study mainly through the learners' learning behavior and learning style and data mining organization personalized learning content, but there are many factors affecting learning behavior, such as the quantitative elements, emotion and attitude, cognitive status, learning motivation, learning strategy, so need to consider other factors affecting the learning behavior and added to the personalized learning model.

References

- [1] Ezaldeen H, Misra R, Bisoy S K, et al. A hybrid E-learning recommendation integrating adaptive profiling and sentiment analysis[J]. *Journal of Web Semantics*, 2022, 72: 100700.
- [2] Promrue S, Sanrattana W. Online Program to Empower Teacher Learning to Develop Students' Digital Literacy Skills[J]. *Education Quarterly Reviews*, 2022, 5(2): 469-483.
- [3] Shemshack A, Spector J M. A comprehensive analysis of personalized learning components [J]. *Journal of Computers in Education*, 2021, 8(4): 485-503.
- [4] Korableva O, Durand T, Kalimullina O, et al. Studying user satisfaction with the MOOC platform interfaces using the example of coursera and open education platforms[C]// *Proceedings of the 2019 International Conference on Big Data and Education*. 2019: 26-30.
- [5] Lee G, Keum S, Kim M, et al. A study on the development of a MOOC design model[J]. *Educational technology international*, 2016, 17(1): 1-37.
- [6] Massive open online course [EB/OL]. [2013-07-22]. <http://en.wikipedia.org/wiki/Massiveopenonlinecourse>.
- [7] Luo Y, Zhou G, Li J. Comparing the chinese university mooc platform to the three major mooc players[C]// *Proceedings of the 2nd International Conference on Computer Science and Application Engineering*. 2018: 1-5.
- [8] Marques J. A short history of MOOCs and distance learning[J]. *MOOC News and Reviews*, 2013, 17.
- [9] Duan Y. Research on joint cultivation mechanism of adult education under collaborative innovation environment in China[J]. *Creative Education*, 2020, 11(05): 797.
- [10] Hummel K A, Hlavacs H. Anytime, anywhere learning behavior using a web-based platform for a university lecture[C]// *Proceedings of the SSGRR 2003 Winter Conference, L'Aquila, Italy*. 2003.
- [11] Graf S. Adaptivity in learning management systems focussing on learning styles[J]. 2007.
- [12] Ribarsky W, Wang D X, Dou W. Social media analytics for competitive advantage[J]. *Computers & Graphics*, 2014, 38: 328-331.
- [13] Saringat M Z B. A Systematic Literature Review Enhanced Felder Silverman Learning Style Models (FSLSM)[C]// *2022 Seventh International Conference on Informatics and Computing (ICIC)*. IEEE, 2022: 1-7.
- [14] Wang J, Mendori T. The reliability and validity of felder-silverman index of learning styles in mandarin version[J]. *Information Engineering Express*, 2015, 1(3): 1-8.
- [15] Rajkumar R, Ganapathy V. Bio-inspired learning style chatbot inventory using brain computing interface to increase the efficiency of e-learning[J]. *IEEE Access*, 2020, 8: 67377-67395.
- [16] Liu S, Lu Y, Zheng P, et al. Adaptive reconstruction of digital twins for machining systems: A transfer learning approach[J]. *Robotics and Computer-Integrated Manufacturing*, 2022, 78: 102390.