Online-Journals.org

iJET International Journal of Emerging Technologies in Learning

iJET | eISSN: 1863-0383 | Vol. 18 No. 13 (2023) | OPEN ACCESS

https://doi.org/10.3991/ijet.v18i13.41909

PAPER

A New Methodology for Clustering of Online Learning Resources Based on Students' Learning Styles

Qiaolan Li(✉)

School of Chinese Language and Culture, Zhejiang Yuexiu University, Shaoxing, China

20192105@zyufl.edu.cn

ABSTRACT

Recommending learning resources to students according to their respective learning style is conductive to improving learning efficiency, and the clustering and classification of learning resources conduce to reducing learning resource redundancy and duplication and increasing learning resource utilization. However, available methods of learning resource recommendation usually regard students' learning styles as fixed and invariable, which apparently contradicts reality and may have a negative influence on students' learning effect. In view of this matter, this paper aims to propose a novel methodology for clustering online learning resources based on student learning style. At first, the specific steps of the new clustering method were given, and a Sharpe model was adopted to analyze the invalid exposure of students' learning effect and identify students' learning styles. The learning style coefficient of students was regarded as a dynamic systemic state, which was estimated by the Kalman filter. Then, the Affinity Propagation Clustering (APC) algorithm was adopted to cluster learning resources based on a student learning style distance matrix, and a model of recommended online learning resource combinations was established based on the proposed method. At last, experimental procedures, including learning style evaluation, pretest exam score prediction, posttest exam score prediction, and data analysis, were described in detail, and the validity of the proposed method was verified by experimental results.

KEYWORDS

learning style, online learning resource, resource clustering

1 INTRODUCTION

Thanks to the development of Internet technology, online learning resources have been greatly enriched and now they are a crucial part of contemporary education [1–4]. Via Internet, students can access massive learning resources, such as course videos, learning materials, exercise questions, and online tutorials [5], however, it is a difficult job for students to choose the right learning resources for themselves

Li, Q. (2023). A New Methodology for Clustering of Online Learning Resources Based on Students' Learning Styles. *International Journal of Emerging Technologies in Learning (iJET)*, 18(13), pp. 234–249. https://doi.org/10.3991/ijet.v18i13.41909

Article submitted 2023-05-01. Resubmitted 2023-06-01. Final acceptance 2023-06-02. Final version published as submitted by the authors.

from such massive amounts of learning resources [6–9]. For the different-type online learning resources, students' acceptance and utilization vary from person to person, and this has posed a great challenge for teachers [10–14]. To better meet the individual needs of each student and help them improve learning effect, it's necessary to analyze their learning styles and cluster those online learning resources with similar styles together, in this way, students could easily find out resources that suit themselves better.

Understanding students' learning styles can help teachers provide personalized learning experiences to students [15–18]. Through clustering methods, students can choose courses and resources that fit their preferences and ability levels. Recommending learning resources to students according to their respective learning style is conductive to improving learning efficiency [19], and the clustering and classification of learning resources also conduce to reducing learning resource redundancy and duplication and increasing learning resource utilization. Hence, researching the online learning resource clustering methods based on student learning style is an important and meaningful work.

Cohn and Holm [20] pointed out that unsupervised machine learning offers an opportunity to extract knowledge from unlabeled data sets and maximize the performance of machine learning, the authors demonstrated in their paper how to construct, use, and evaluate a high-performance unsupervised machine learning system for classifying images in a popular micro-structural dataset. After signals were extracted from feature descriptors by principal component analysis, k-means clustering was adopted for image analysis without the necessity of labeling the training data, and the proposed method achieved an accuracy of 99.4% ± 0.16%. Li et al. [21] proposed to perform online clustering by conducting twin contrastive learning (TCL) at the instance level and the cluster level, the authors discovered that when data is projected into a feature space with a dimensionality of the target cluster number, the rows and columns of its feature matrix correspond to the instance and cluster representation, respectively, and the proposed TCL constructed positive and negative pairs for a given dataset through data augmentations based on the observations. Jing Dong et al. [22] argue that with the development of Internet+education, learners using online learning platform to carry out independent learning has entered a popularization phase, the authors constructed an online learning behavior analysis model and applied a clustering algorithm based on data mining technology to analyze and generate learning behavior analysis reports, thereby intervening the teaching, urging teachers to improve their teaching quality, and providing suggestions for learners to improve their learning behavior.

After carefully reviewing relevant literatures, it's found that available methods of resource clustering have seldom been applied to the clustering of online learning resources, few attention has been paid to the analysis of students' learning styles, and the existing learning resource recommendation usually regards students' learning styles as fixed and invariable. However, in fact, students' learning styles may change with time, course content, and teaching method, so the neglect of the dynamic changes of students' learning styles may have a negative influence on students' learning effect. Existing resource clustering methods mostly rely on average or overall data, they haven't fully considered the uniqueness of each student, but actually each student has his or her unique style in learning and they vary in learning ability as well. Thus, a method applicable to most students may not suit all of them. In view of this matter, this paper studied a novel methodology for clustering online learning resources based on student learning style.

2 ANALYSIS OF STUDENT LEARNING STYLES

Specific steps of online learning resource clustering based on student learning style involve four aspects: first, collect data of students' learning situations and online learning resources, including their learning behavior, academic performance, and learning time, etc.; in the meantime, data of online learning resources needs to be collected as well, such as teaching videos, handouts, and exercises, etc. Second, further analyze students' learning styles, which can be defined from multiple dimensions such as the students' behaviors, preferences, and abilities. Third, the online learning resources are classified according to learning styles, for example, for visual learners, resources containing many images and videos can be gathered together; while for theoretical learners, resources containing various theoretical interpretations can be collected for them. At last, give personalized online learning resource recommendations to help students find the most suitable learning resources and improve learning efficiency and effectiveness. Figure 1 shows the research path of this paper.

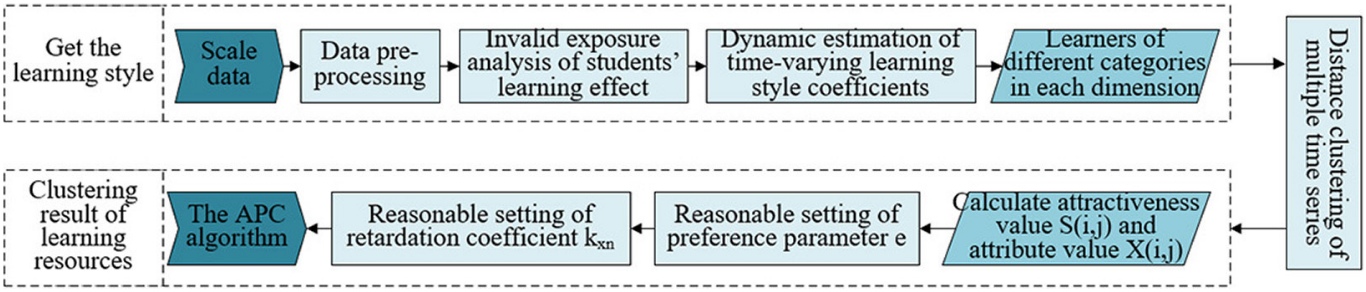

Fig. 1. Research path of this paper

Sharpe model, namely the knowledge exposure model, is a tool for evaluating the improvement of learning effect. This model views the learning process as an accumulation of a series of exposures, and each exposure represents an opportunity for learners to contact a new concept. In this paper, a valid exposure is defined as that a learner has contacted a new concept and can understand and remember it; an invalid exposure is defined as that a learner has contacted a new concept but can not understand or remember it. By combining these definitions and the identification of students' learning styles, we can better comprehend and analyze students' learning effect. Specifically, learning style identification can help teachers figure out students' learning habits and advantages, so that the teachers can design teaching methods that suit the students better, thereby increasing the opportunity of valid exposure. This paper used the *Sharpe* model to analyze the invalid exposure of students' learning effect and realize the identification of students' learning styles.

Sharpe model performs regressions on learning effect improvement rate sequence and style coefficient improvement rate sequence of students within a certain time period in the past, and splits the learning effect combination improvement value of students into different invalid learning resource items, thereby attaining the learning style coefficient of each item. Assuming: $S_{i,o}$ represents the improvement rate of student learning effect i at time moment o, L represents the number of style coefficients, $G_{l,o}$ represents the improvement rate of style coefficient l at time moment o, $y_{i,l}$ represents the coefficient of student learning effect improvement rate with respect to style coefficient improvement rate, $p_{l,o}$ represents the residual of student learning effect i at time moment o, then there are:

$$S_{i,o} = y_{i,1}G_{1,o} + y_{i,2}G_{2,o} + \cdots + y_{i,L}G_{L,o} + p_{i,o}$$
$$s.t.\begin{cases} y_{i,1} + y_{i,2} + \cdots + y_{i,L} = 1 \\ y_{i,1}, y_{i,2}, \ldots, y_{i,l} \geq 0 \end{cases} \tag{1}$$

The model has both equality constraints and inequality constraints, thus it needs be converted into a quadratic programming problem with constraint conditions before processed by the least squares method. According to variance of maximized S^2 and minimized $p_{i,o}$, an objective function was set as follows:

$$\min(\mathrm{var}(p_i)) = \min(P(p_i^2) - (P(p_i))^2)$$
$$s.t.\begin{cases} p_i = S_i - (y_{i1}G_1 + y_{i2}G_2 + \cdots + y_{il}G_l) \\ \sum_{j=1}^{L} y_{il} = 1 \\ 0 \leq y_{il} \leq 1 \\ l = 1, 2, \ldots, L \end{cases} \tag{2}$$

Kalman filter is an efficient recursive filter that can estimate the state of linear dynamic systems in case there are uncertainties. In terms of the analysis of students' learning effect, the Kalman filter can make timely adjustments to the estimated results with its dynamic adaptability when the learning style coefficient of students undergoes changes, and this is quite useful for the dynamic tracking and analysis of students' learning styles. In addition, Kalman filter can make effective use of history data and current observation data to reduce estimation errors, which can improve the estimation accuracy of student learning style coefficient. In this way, the Kalman filter can give valid state estimations in the presence of data uncertainties, which is very useful for processing students' learning data, because students' learning behavior and performance are affected by many uncertain factors. Due to these reasons, this paper took the student learning style coefficient as the state of a dynamic system, and adopted the Kalman filter to perform dynamic estimation on it.

Assuming that the learning style of students' learning effect is time-varying, then there are:

$$S_o = y_{1,o}G_{1,o} + y_{2,o}G_{2,o} + \cdots + y_{L,o}G_{L,o} + p_o \tag{3}$$

$$y_{l,o+1} = y_{l,o} + \delta_{l,o+1} \tag{4}$$

Assuming Ω is an independent normally distributed random sequence, then the error term of above formula satisfies: $p_o \sim \Omega(0, \varepsilon_p^2)$, $\delta_{l,o} \sim \Omega(0, \varepsilon_{l,n}^2)$.

Through above steps, it's known that the model allows the style coefficient to adapt to the changes in learning style since the style coefficient in the model is not constant.

The constructed model is a state space model composed of measurement equations and state equations, and the parameters of the model can be directly solved by the Kalman filter. Assuming: b_o represents the observed variable at time moment o, η_o represents the state variable at time moment o, then the relationship between b_o and η_o can be represented by an observation matrix F, the relationship between the η_o at current moment and the η_o at previous moment can be represented by a transition matrix X, q_o and u_o are error terms of the two equations, then the standard form of the constructed model is given by following formulas:

$$b_o = F \cdot \eta_o + q_o \tag{5}$$

$$\eta_o = X \cdot \eta_{o-1} + u_o \tag{6}$$

$$P(q_o q_\kappa^{'}) = \begin{cases} J, o = \kappa \\ 0, o \neq \kappa \end{cases} \tag{7}$$

$$P(u_p u_\kappa^{'}) = \begin{cases} L, o = \kappa \\ 0, o \neq \kappa \end{cases} \tag{8}$$

Corresponding to the standard form of the state space model, the time-varying learning style *Sharpe* model of $b_o = S_o$, $\eta_o = y_o$, $F = G$, $X = 1$, $q_o = p_o$, $u_o = \delta_o$ could be set; for this state space model, the recursive process of the Kalman filtering method has these steps:

Step 1: Assuming $\eta^{\wedge}_{o|o-1} = P(\eta_o) = X\eta^{\wedge}_{o|o-1}$ represents the expected value of state variable after state transition, calculate $\eta^{\wedge}_{o|o-1}$. In case there's no b_o, calculate the current expected value based on the expected value of the previous moment, wherein the initial value $\eta^{\wedge}_{1|0}$ is known and it satisfies $\eta^{\wedge}_{1|0} \sim N(0, \varepsilon^2_0)$, it's assumed that the mean square error of $\eta^{\wedge}_{o|o-1}$ is represented by $E_{o|o-1} = P[(\eta_o - \eta^{\wedge}_{o|o-1})(\eta_o - \eta^{\wedge}_{o|o-1})'] = XE_{o-1}X'+K$.

Step 2: Calculate the expected value of the observed variable $b^{\wedge}_{o|o-1} = P(b_o) = F\eta^{\wedge}_{o|o-1}$.

Step 3: Assuming L represents the Kalman gain matrix, based on the actual value of b_o, update the expected value of η_o, $\eta^{\wedge}_o = P(\eta_o + L(b_o - b^{\wedge}_{o|o-1})) = \eta^{\wedge}_{o|o-1} + L(b_o - F\eta^{\wedge}_{o|o-1})$, wherein L satisfies $L = E_{o|o-1}F' (FE_{o|o-1}F'+J)^{-1}$.

Step 4: Calculate the mean square error $E_o = (1-LH)E_{o|o-1}$ after the update.

By substituting the $\eta^{\wedge}_o$ attained in Step 3 and the E_o attained in Step 4 into the Step 1 of the next time moment $o+1$, proceed to the next Kalman cycle.

In this paper, the parameters of Kalman filtering process were solved by the maximum likelihood estimation method. Assuming: η_o, q_o, and u_o obey normal distribution, then the conditional distribution of variable b_o also conforms to normal distribution, the mean value is $P(b_o) = F\eta^{\wedge}_{o|o-1}$, the variance is $P[(b_o - b^{\wedge}_{o|o-1})(b_o - b^{\wedge}_{o|o-1})'] = FE_{o|o-1}F'+J$, and the maximum likelihood function is $g(b_o, B_{o-1}) = (2\pi)^{-\pi/2}(FE_{o|o-1}F'+J)^{-1/2} \times exp[-1/2 (b_o - F\eta^{\wedge}_{o|o-1})(FE_{o|o-1}F'+J)^{-1}(b_o - F\eta^{\wedge}_{o|o-1})']$, $o = 1,2,\ldots,O$.

For the time-varying *Sharpe* model constructed without constraints, the sequence of dynamic style coefficients could be attained through above steps, while for models with equality constraints and without inequality constraints, the linear transformation method could be adopted for processing. $S_o = \sum_{l=1}^{L} y_{l,o}G_l + p_o$ can be expressed as $S_o = \sum_{l=1}^{L-1} y_{l,o}G_1 + (1 - \sum_{l=1}^{L-1} y_{1,o})Gl + p_o$, that is $(S_o - G_L) = \sum_{l=1}^{L-1} y_{l,o}(G_l - G_L) + p_o$; by performing Kelman filtering on the observed value sequence $(S_o - G_L)$ and each student online learning style $(G_l - G_L)$, the dynamic multi-variate style coefficients of the model with equality constraints and without inequality constraints can be attained.

3 CLUSTERING OF LEARNING RESOURCES BASED ON STUDENT LEARNING STYLE

After obtaining the distance matrix between learning resources, clustering algorithms can be adopted to cluster resources in the online learning resource pool. The learning resources in each class are resources with high correlation, the class center represents the most representative learning resource in this class. By selecting the cluster center of each resource class in the resource pool, the correlation of learning resources in the learning resource combinations recommended to students could be decreased effectively.

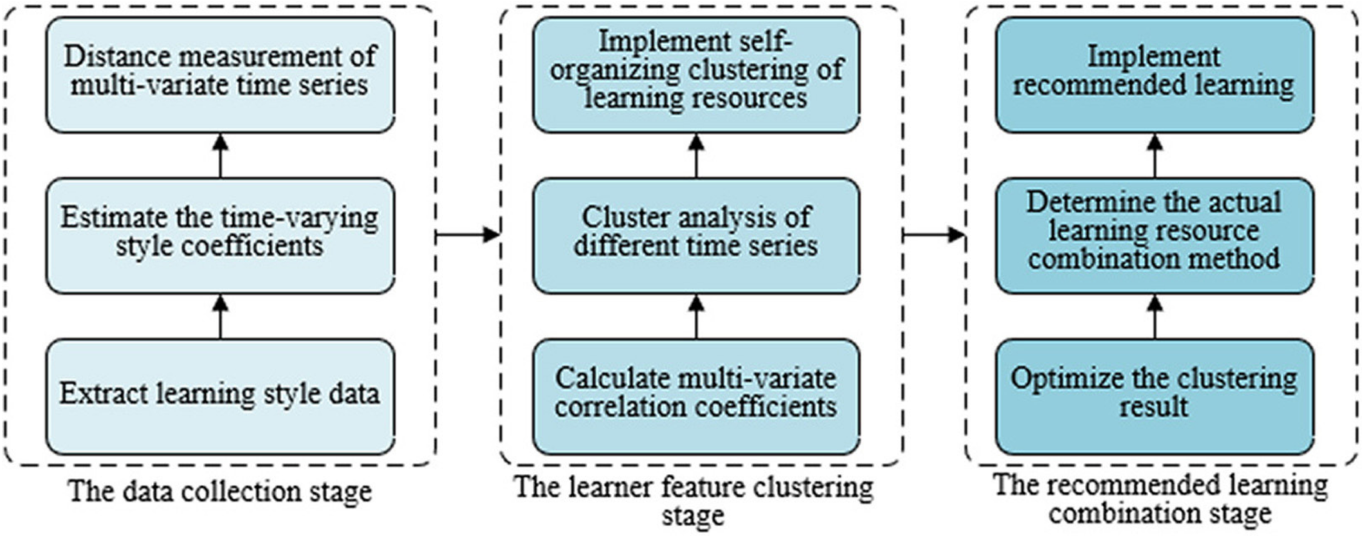

Fig. 2. The model of recommended online learning resource combinations established based on the proposed method

APC is a clustering method that does not require to set the cluster number in advance. Conventional clustering methods such as *K-means* generally require to set the cluster number in advance, while APC doesn't need it. APC can automatically determine the optimal cluster number according to data, which makes it advantageous in clustering online learning resources since the diversity and dynamics of resources make it very difficult to set the cluster number in advance. APC takes the similarity between all data points into consideration, so it can cluster more accurately. In the scenario of this study, it means that APC can classify learning resources based on the similarity of student learning styles more accurately, so it has been chosen in this paper and used to cluster learning resources based on the distance matrix of student learning styles. Figure 2 shows the model of recommended online learning resource combinations established based on the proposed method.

In this study, the online learning resources were clustered into 10 classes based on learning style analysis: 1) Visual learning resources: including figures, tables, images, presentations and other learning resources presented in the visual form; 2) Audio learning resources: including audio files, podcasts, recordings and other learning resources presented in the sound form; 3) Text learning resources: including books, articles, manuals and other learning resources presented in the written form; 4) Interactive learning resources: including simulators, virtual labs, online courses and other learning resources that require students to participate in interactions; 5) Social learning resources: including online discussions, groups, collaborations, and other learning resources that require students to interact with others; 6) Practical learning resources: including field visits, internships, practical projects and other learning resources that require students to carry out practical operations; 7) Analytical learning resources: including data analysis, statistical analysis, case analysis and other learning resources that require students to analyze and reason; 8) Theoretical learning resources: including academic papers, theoretical frameworks, conceptual models and other learning resources that require students to think and discuss; 9) Applied learning resources: including practical application cases, industry practices, practical skills and other learning resources that require students to learn practical applications and operation skills; 10) Mixed learning resources: including the combinations of multiple learning resources, such as video+interaction, and text+practice, etc.

Before clustering, calculating the distance of multi-variate time series of student learning styles is a very important work, through which the similarity of learning styles between students can be identified and quantified, and this is critical for subsequent clustering and learning resource recommendation, because in this way students with similar learning styles can be grouped into a same class and similar learning resources can be recommended to them. Moreover, students' learning styles may change over time, by calculating the distance of multi-variate time series, we can also track these changes and adjust the clustering and recommendations of learning resources in time.

For the series of multi-variate style coefficients obtained based on the regression of w classes of student learning style coefficients in the previous chapter, a time window length O was set, then the size of the data matrix of each learning resource is $O \times w$. Assuming: A and B are the time series data of two different learning resources, AA' represents the inner product of matrix A, $tr(\ )$ represents the trace operation of the matrix, then the multi-variate correlation coefficient between the two can be attained from the following formula:

$$SU_{AB}^2 = \frac{(AA', BB')}{\|AA'\|\|BB'\|} = \frac{tr(AA', BB')}{\sqrt{tr(AA'^2)tr(BB'^2)}} = \frac{tr(R_{AB}, R_{BA})}{\sqrt{tr(R_{AB}^2)tr(R_{BB}^2)}} \tag{9}$$

The covariance matrix of sample A can be calculated using the following formula:

$$R_{AA} = \frac{1}{O-1} AA' \tag{10}$$

The covariance matrix between A and B can be attained through the following formula:

$$R_{AB} = \frac{1}{O-1} AB' \tag{11}$$

To enable the correlation coefficients directly used by the clustering algorithm, they were transformed as follows:

$$c_{AB} = \sqrt{2(1 - SU_{AB})} \tag{12}$$

In this study, the APC algorithm was chosen and used to cluster the learning resources based on the distance matrix of student learning styles. Unlike other clustering algorithms, such as *k-means* which requires to set the number of clusters in advance, APC can find the optimal cluster number automatically. During the application of teaching resource clustering, the optimal cluster number is usually unclear to us, so this advantage is particularly important. The APC algorithm considers the contribution of all data points to cluster representatives, not just the cluster center or some certain points, and this enables it to get better results when dealing with complex-structured data, such as the multi-dimensional data of student learning styles.

APC algorithm sets two indexes attractiveness value $S(i,j)$ and attribute value $X(i,j)$ to describe the relationship between sample points, wherein $S(i,j)$ represents the attractiveness of sample i for sample k, $X(i,j)$ represents the attribution of sample i to sample l; it also sets a preference parameter e to control the difficulty for each sample to become the cluster center, then there are:

$$S(i,l) = R(i,l) - \max_{j \ s.t. j \neq l}(X(i,j) + R(i,j)) \tag{13}$$

$$S(l,l) = e - \max_{j}(X(l,j) + R(l,j)) \tag{14}$$

$$X(i,l) = \min\left[S(l,l) + \sum\nolimits_{j \ s.t. j \neq i,l} \max(S(j,l),0),0\right] \tag{15}$$

$$X(l,l) = \sum\nolimits_{j \ s.t. j \neq i,l} \max(S(j,l),0) \tag{16}$$

It can be known that, if e is small, then $S(l,l)$ and $X(i,l)$ will decrease accordingly, and the probability of a sample becoming the cluster center will decrease as well, so the class number of learning resource clusters can be controlled by the value of e.

Besides, the retardation parameter k_{xn} that determines the update amplitude between samples $S(l,l)$ and $X(i,l)$ during each iteration needs to be set reasonably:

$$S_{new}(i,l) = (1 - k_{xn})S(i,l) + k_{xn}S_{old}(i,l) \tag{17}$$

$$X_{new}(i,l) = (1 - k_{xn})X(i,l) + k_{xn}X_{old}(i,l) \tag{18}$$

k_{xn} can affect the update speed of the model. If its value is set too large, then the algorithm may run too slowly; if it its value is set too small, the clustering result may be unstable and oscillate back and forth. Thus, in order to get a clustering result of the required number of online learning resource classes, e needs to be assigned with a relatively large value at first and then be deceased gradually, in this way, the number of online learning resource clusters reduces gradually until the required clustering result of the number of online learning resource classes is attained.

4 EXPERIMENTAL RESULTS AND ANALYSIS

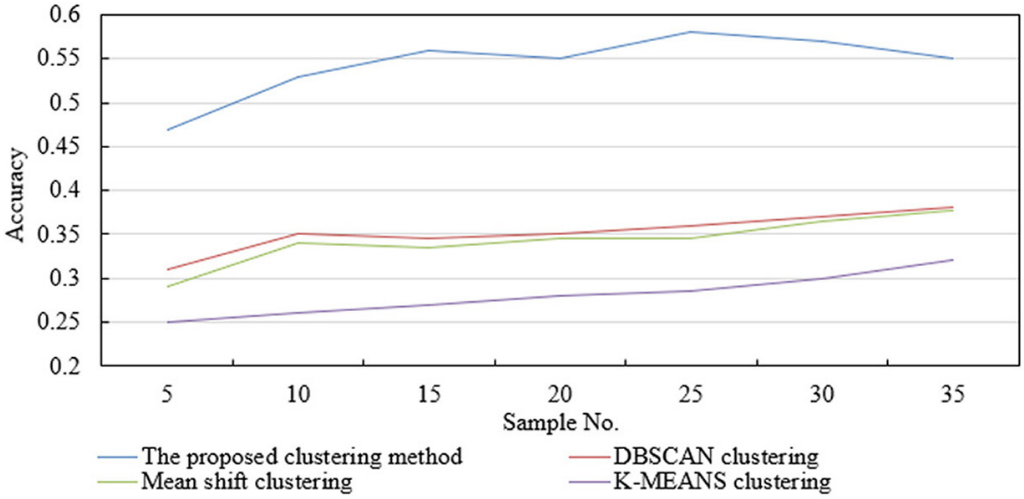

Fig. 3. Accuracy comparison of different clustering algorithms

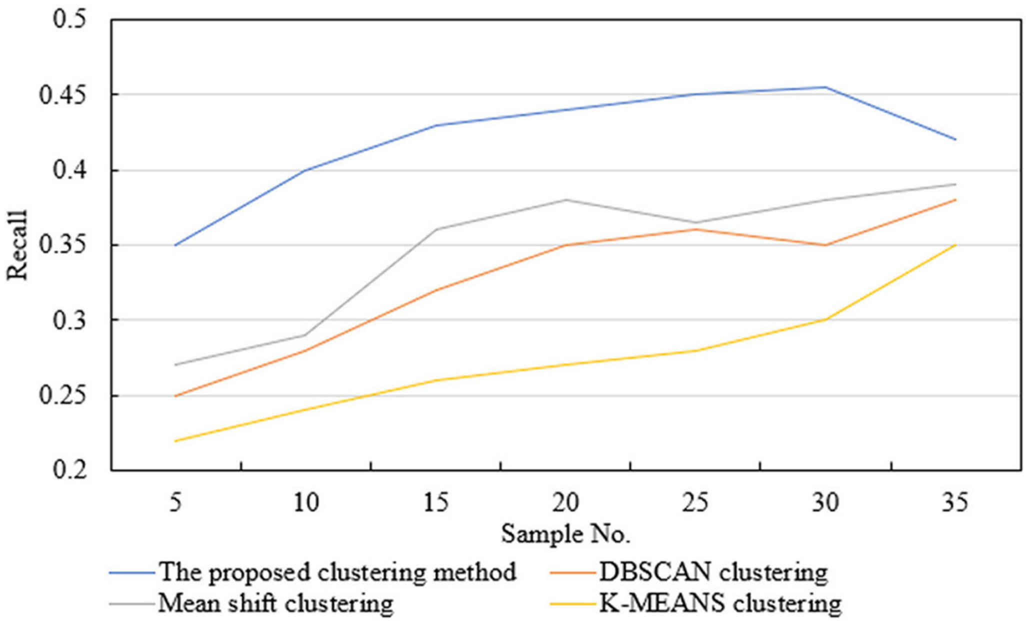

Fig. 4. Recall rate comparison of different clustering algorithms

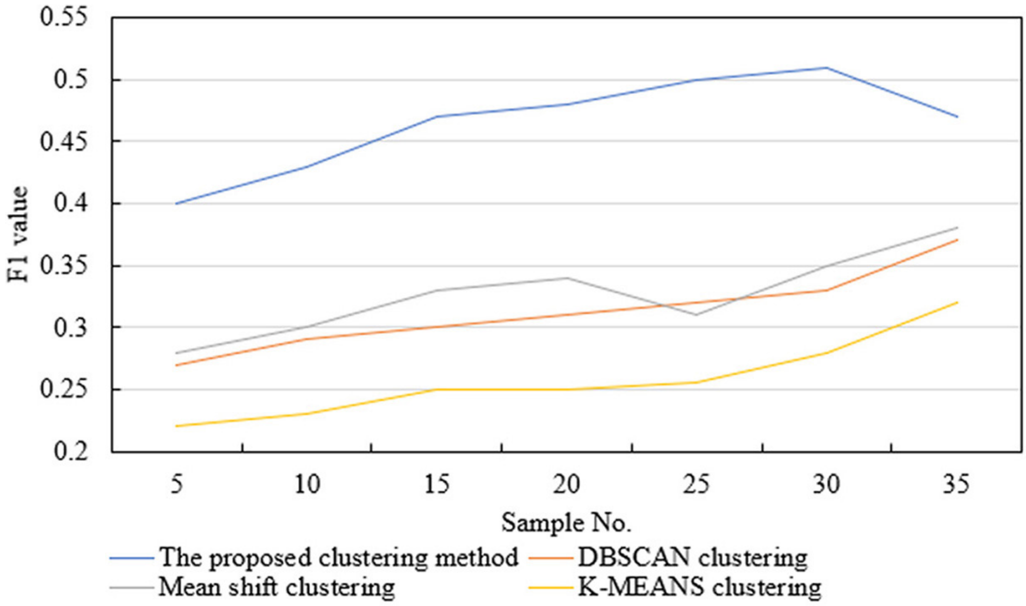

Fig. 5. F1 value comparison of different clustering algorithms

Figures 3–5 respectively compare the accuracy, recall rate, and F1 value of different clustering algorithms. According to the figures, under the condition of all sample sizes, the proposed clustering method (namely the APC algorithm based on student learning style) exhibited high accuracy, recall rate, and F1 value. In contrast, the accuracy, recall rate, and F1 value of DBSCAN clustering, mean shift clustering, and K-MEANS clustering were lower.

By analyzing the reasons, it's known that the APC method has a few merits in clustering, including automatically determining the cluster number, global optimal solution, dynamic adaptability and robustness to abnormal values, which all contribute to the improvement of clustering accuracy to some extents. In contrast, DBSCAN clustering can effectively deal with noise and outliers, but it's sensitive to parameter setting; mean shift clustering can determine cluster number automatically, but it is sensitive to initial setting and only has a low efficiency in processing large-scale data; K-MEANS clustering is easy in calculation but it needs to set cluster number in advance and can only attain local optimal solution. In summary, the proposed method outperformed DBSCAN clustering, mean shift clustering, and K-MEANS clustering out of its advantages in automatically determining cluster number, searching for global optimal solution, dynamic adaptability and robustness to outliers, thereby, using it to cluster online learning resources can effectively improve the accuracy and efficiency of learning resource recommendation.

Table 1 lists the learning styles of different students in four aspects: personalized learning, classroom learning assistance, knowledge expansion and innovation, and technology capability upgrade. In this paper, these four aspects were taken as the four evaluation dimensions of students' learning styles, and the learning style in each dimension was attained by observing and analyzing the actual learning behavior of students. According to the table, students' learning styles in the aspect of personalized learning can be divided into two types: cooperative and independent. Some students are better at cooperation and enjoy the interactive and sharing experiences brought by group learning, and such learning style is very beneficial to cultivating students' team work ability and social communication skills. Some students are better at learning independently, and they prefer to solve problems through their own efforts, and such learning style is conductive to cultivating students' autonomous learning ability and problem solving ability. In terms of classroom learning assistance, students' learning styles can be divided into the comprehension type and the practice type. The comprehension-type students understand the inherent logic and laws of the knowledge first, and then apply the knowledge to practice; while practice-type students are more inclined to understand and master knowledge through practical activities. In terms of knowledge expansion and innovation, students' learning styles can be divided into two types: interactive and explorative. Students of the first type prefer to expand their knowledge and acquire new ideas through communications and interactions with people; while students of the second type prefer to obtain new knowledge and ideas through their own thinking and exploration. In terms of technology capability upgrade, students' learning styles can be divided into analysis type and verification type. Analysis-type students are good at improving their technical ability through analysis and thinking; while verification-type students are more inclined to improve through practice and verification.

Table 1. Identification results of some students' learning styles

Student No.	Personalized Learning	Classroom Learning Assistance	Knowledge Expansion and Innovation	Technology Capacity Upgrade
0	Cooperative	Comprehension	Interactive	Analysis
1	Independent	Practice	Explorative	Analysis
2	Independent	Practice	Explorative	Verification
3	Cooperative	Comprehension	Interactive	Verification
4	Independent	Practice	Interactive	Analysis
5	Cooperative	Practice	Explorative	Verification

(Continued)

Table 1. Identification results of some students' learning styles *(Continued)*

Student No.	Personalized Learning	Classroom Learning Assistance	Knowledge Expansion and Innovation	Technology Capacity Upgrade
6	Cooperative	Comprehension	Interactive	Analysis
7	Independent	Comprehension	Explorative	Verification
8	Cooperative	Practice	Explorative	Verification
9	Cooperative	Comprehension	Interactive	Analysis
10	Cooperative	Comprehension	Interactive	Analysis

Based on the identification results of students' learning styles, the proportions of learning resources of 10 different classes were analyzed, as shown in Figure 6, the distribution of these 10 classes of learning resources is not uniform. Classes 5 and 6 account for the a higher proportion, reaching 15.2% and 14% respectively, this indicates that these two classes of learning resources can meet the learning needs of most students, and they are more universal and suitable for most students. Classes 2 and 4 also account for a relatively large proportion, both exceeding 10 percent, this indicates that these two classes of learning resources are also popular among students, or they target at some common learning tasks or knowledge points. In contrast, Classes 1, 3, and 9 account for a low proportion, all below 10%, indicating that these learning resources are only about some special learning tasks or knowledge points, and their application range is narrow and limited. The proportion of Class 10 is the lowest, only 2%, indicating that these learning resources are very special and apply only to a few students, or their quality and effect need to be improved further. Overall, these numbers have reflected that learning resources are diverse and students' needs are personalized.

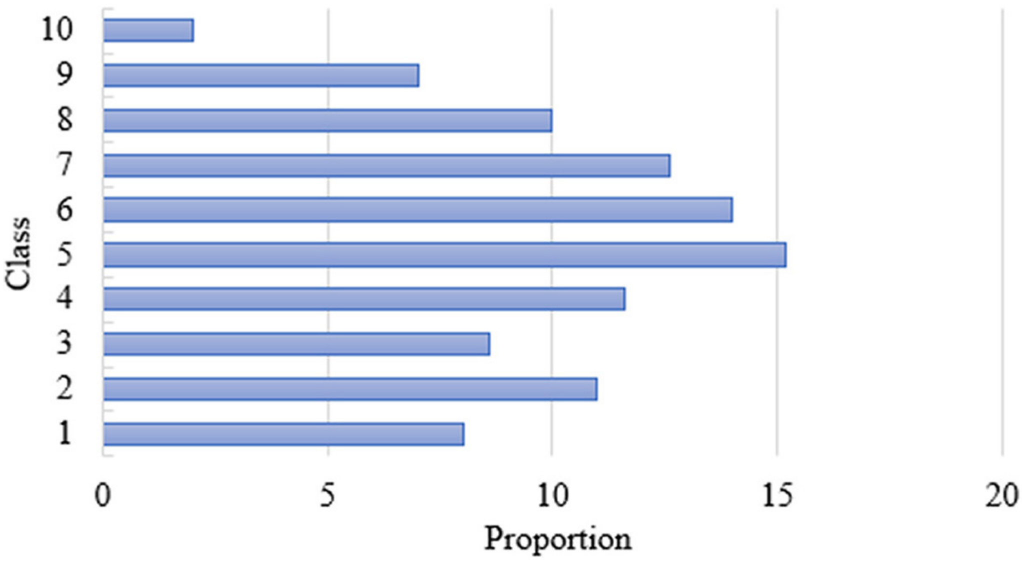

Fig. 6. Proportions of learning resources by class

Table 2. Learning styles of students corresponding to learning resources with the highest and lowest scores

Learning Resource Type	The Highest			
1	Cooperative	Comprehension	Explorative	Analysis
2	Cooperative	Practice	Interactive	Analysis
3	Cooperative	Practice	Interactive	Verification
4	Independent	Comprehension	Explorative	Analysis
5	Cooperative	Comprehension	Interactive	Verification
6	Cooperative	Comprehension	Explorative	Verification/Analysis
7	Independent	Comprehension	Explorative	Verification
8	Independent	Comprehension	Explorative	Verification
9	Independent	Practice	Interactive	Verification
10	Cooperative	Practice	Explorative/Interactive	Analysis
Learning Resource Type	**The Lowest**			
1	Cooperative	Practice	Explorative	Analysis
2	Cooperative	Practice	Interactive	Verification/Analysis
3	Independent	Practice	Interactive	Analysis
4	Independent	Comprehension	Interactive	Verification
5	Independent	Comprehension	Explorative/Interactive	Verification
6	Independent	Comprehension	Interactive	Verification
7	Independent	Practice	Explorative/Interactive	Analysis
8	Cooperative	Practice	Explorative	Verification/Analysis
9	Independent	Practice	Explorative	Verification
10	Cooperative	Practice	Explorative	Verification

Table 2 summarizes the learning styles of students corresponding to learning resources with the highest and lowest scores. According to the table content, it can be seen that different types of learning resources match with different learning styles. For "cooperative" and "comprehension" learning styles, analytical-type learning resources score higher than explorative-type learning resources, this is because in collaborative scenarios, analytical skills are more important in helping students better understand the information, while explorative-type learning resources require a higher degree of autonomy. For "practice" style learners, some resources are suitable for analysis and verification learning styles. In practical learning, students learn new knowledge and skills through hands-on practice, and this requires students to have the analysis and verification ability, so such learning resources got a higher score. For "independent" style learners, comprehension-type and explorative-type resources scored the highest in analysis-style and scored the lowest in verification-style, this is because independent learners like to understand and explore knowledge by themselves, so they do not need too much verification. In the meantime, in some learning resource classes, cooperative and independent learning styles can both find suitable resources, indicating that the online learning resources need to consider diversity and personalized needs.

Table 3 summarizes the changes in scores of learning resources before and after clustering, it can be seen that after clustering, there are significant changes in students' learning resource scores. In terms of resources of Class 1, both the maximum and minimum values increased after clustering, indicating that some efficient learning resources have been successfully recommended to some students through clustering, so students' learning behavior in this category has increased. For resources of Class 2, after clustering, the mean, maximum and minimum values decreased, indicating resources of this class have been inhibited, this is because they might not be the most effective for the learning of most students, and the clustering has found more suitable learning resources for students. For resources of Class 3, after clustering, the maximum and minimum values decreased, indicating that the students' learning behavior has become more concentrated, the clustering has helped them reduce time spent on unnecessary learning. For resources of Class 6, after clustering, the minimum value increased significantly, indicating that the learning behavior of this class is helpful for all students, and the clustering allows all students to access these efficient learning resources. As for resources of Classes 7 and 8, after clustering, the maximum and minimum values increased, indicating that the students' activeness degree of these two classes of resources has increased. In terms of Classes 9 and 10, after clustering, the maximum and minimum values also increased, indicating that the clustering has enabled some students to find learning resources that are more suitable for them, and their learning efficiency has been improved. Based on the above analysis, it can be concluded that, after the clustering of online learning resources based on learning style, the online learning behavior of students has changed significantly, overall speaking, the clustering helps students make better use of the online learning resources and their learning efficiency has been improved.

Table 3. Changes in scores of learning resources before and after clustering

Score Resource No.	Before				After	
	Mean	Standard Deviation	Maximum	Minimum	Maximum	Minimum
1	83.1	33.6	147	23	168	1
2	81.9	80.4	362.8	12.4	73.6	5.2
3	52.6	22.1	141	29	119	9.6
4	9.1	4.2	15	2	7	2
5	83.7	4.6	130	63	114	41.5
6	80.2	2.1	162	69	152	215
7	0.41	2.8	5	2	6	3
8	3.1	7.3	18	4	8	6
9	74.3	6.1	127	65.2	184	39.6
10	81.9	4.3	195	71.4	169	44.1

5 CONCLUSION

This study explored a new clustering method of online learning resources based on students' learning styles. At first, the specific steps of the new clustering method were given, and a *Sharpe* model was adopted to analyze the invalid exposure of students' learning effect and identify students' learning styles. The learning style coefficient of students was regarded as a dynamic systemic state, which was estimated by the Kalman filter. Then, the APC algorithm was adopted to cluster learning resources based on the student learning style distance matrix, and a model of recommended online learning resource combinations was established based on the proposed method. In the experimental part, the specific procedures of the experiment, including the evaluation of learning styles, statistics of pretest exam scores, statistics of posttest exam scores, and data analysis, were introduced in detail; the accuracy, recall rate and F1 value of different clustering algorithms were compared; and the learning styles of different students in four aspects of personalized learning, classroom learning assistance, knowledge expansion and innovation, and technology capability upgrade were listed. After that, based on the identification results of students' learning styles, the proportions of learning resources in 10 classes were analyzed further, and the learning styles of students corresponding to the learning resources with the highest and lowest scores were summarized. At last, the changes in scores of learning resources before and after clustering were given, and the results have verified that after the clustering of online learning resources based learning style has been implemented, students' online learning behavior has undergone significant changes.

6 REFERENCES

[1] Ouyang, S. (2023). Personalized combination recommendation of short video online learning resources. International Journal of Emerging Technologies in Learning, 18(6): 162–176. https://doi.org/10.3991/ijet.v18i06.38719

[2] Arya, D.M., Guo, J.L., Robillard, M.P. (2023). How programmers find online learning resources. Empirical Software Engineering, 28(2): 23. https://doi.org/10.1007/s10664-022-10246-y

[3] Li, J., Lin, N. (2022). Network optimization of online learning resources from the perspective of knowledge flow. International Journal of Emerging Technologies in Learning, 17(16): 137–149. https://doi.org/10.3991/ijet.v17i16.33765

[4] Wen, J., Wei, X.C., He, T., Zhang, S.S. (2020). Regression analysis on the influencing factors of the acceptance of online education platform among college students. Ingénierie des Systèmes d'Information, 25(5): 595–600. https://doi.org/10.18280/isi.250506

[5] Seryakova, S.B., Zhang, X., Galustyan, O.V., Askhadullina, N.N., Pushkareva, T.V., Zvonova, E.V. (2022). Application of online learning within education of future engineers during the Covid-19 pandemic. International Journal of Engineering Pedagogy, 12(1): 95–103. https://doi.org/10.3991/ijep.v12i1.25009

[6] Toporkov, V., Yemelyanov, D., Bulkhak, A. (2022). Machine learning for job-flow online scheduling and resources allocation in distributed computing. In 2022 VI International Conference on Information Technologies in Engineering Education (Inforino), Moscow, Russian Federation, pp. 1–5. https://doi.org/10.1109/Inforino53888.2022.9782965

[7] Afifi, S., Santoso, H.B., Hasani, L.M. (2022). Investigating students' online self-regulated learning skills and their e-learning experience in a prophetic communication course. Ingénierie des Systèmes d'Information, 27(3): 387–397. https://doi.org/10.18280/isi.270304

[8] Ashar, M., Kamdi, W., Kurniawan, D.T. (2021). Professional skills development through the network learning community using an online learning platform. International Journal of Interactive Mobile Technologies, 15(12): 202–210. https://doi.org/10.3991/ijim.v15i12.21587

[9] Wang, S.Y. (2021). Online learning behavior analysis based on image emotion recognition. Traitement du Signal, 38(3): 865–873. https://doi.org/10.18280/ts.380333

[10] Nonthamand, N. (2023). Development of an online learning lesson in the educational technology equipment operation course for pre-service teachers. International Journal of Interactive Mobile Technologies, 17(10): 30–46. https://doi.org/10.3991/ijim.v17i10.37775

[11] Cai, W., Peng, Y. (2022). Sharing method of online teaching resources of spoken english based on deep learning. Journal of Electrical and Computer Engineering, 2022: 9520630. https://doi.org/10.1155/2022/9520630

[12] Ivan Stevanus, Maria Gita, Debor. (2022). The impact of online learning toward the independence of primary school students during the pandemic in Jakarta, Indonesia. International Journal of Online and Biomedical Engineering, 18(8): 23–33. https://doi.org/10.3991/ijoe.v18i08.31965

[13] Liu, Y., Ji, M., Lin, S.S., Zhao, M., Lyv, Z. (2021). Combining readability formulas and machine learning for reader-oriented evaluation of online health resources. IEEE Access, 9: 67610–67619. https://doi.org/10.1109/ACCESS.2021.3077073

[14] Jalinus, N., Ganefri, Zaus, M.A., Wulansari, R.E., Nabawi, R.A., Hidayat, H. (2022). Hybrid and collaborative networks approach: Online learning integrated project and Kolb learning style in mechanical engineering courses. International Journal of Online and Biomedical Engineering, 18(15): 4–16. https://doi.org/10.3991/ijoe.v18i15.34333

[15] Tseng, F.C., Liu, P.H.E. (2021). The development of effective evaluation criteria for online English learning resources. In 2021 International Conference on Advanced Learning Technologies (ICALT), Tartu, Estonia, pp. 279–281. https://doi.org/10.1109/ICALT52272.2021.00090

[16] Pradipta, R.F., Hermanto, Y.A.L., Oktaviani, H.I., Dewantoro, D.A., Yasin, M.H.M. (2021). Development of online learning resources as a support system learning from home at SLB LAB UM. In 2021 7th International Conference on Education and Technology (ICET), Malang, Indonesia, pp. 100–106. https://doi.org/10.1109/ICET53279.2021.9575066

[17] Shi, Y., Pu, Q., Chen, L., Jiang, Y., Yang, H.H. (2021). Research on influencing factors of elementary school students' perceptions of online learning resources during the COVID-19 pandemic. In 2021 Tenth International Conference of Educational Innovation through Technology (EITT), Chongqing, China, pp. 117–122. https://doi.org/10.1109/EITT53287.2021.00031

[18] Pan, X., Song, J., Zhao, J., Truong, V.A. (2020). Online contextual learning with perishable resources allocation. Iise Transactions, 52(12): 1343–1357. https://doi.org/10.1080/24725854.2020.1752958

[19] Utunen, H., Attias, M., George, R., Ndiaye, N., Piroux, C., Farzi, M.R., Sy, A., Gamhewage, G. (2020). Global access to OpenWHO's online learning resources for COVID-19. Studies in Health Technology and Informatics, 272: 304–305. https://doi.org/10.3233/shti200555

[20] Cohn, R., Holm, E. (2021). Unsupervised machine learning via transfer learning and k-means clustering to classify materials image data. Integrating Materials and Manufacturing Innovation, 10(2): 231–244. https://doi.org/10.1007/s40192-021-00205-8

[21] Li, Y., Yang, M., Peng, D., Li, T., Huang, J., Peng, X. (2022). Twin contrastive learning for online clustering. International Journal of Computer Vision, 130(9): 2205–2221. https://doi.org/10.1007/s11263-022-01639-z

[22] Dong, J., Wang, Z., Li, S. (2022). Analysis of online learning behavior based on clustering algorithm. In 2022 IEEE 2nd International Conference on Power, Electronics and Computer Applications (ICPECA), Shenyang, China, pp. 364–368. https://doi.org/10.1109/ICPECA53709.2022.9718935

7 AUTHOR

Qiaolan Li, PhD from Shandong University, professor of Zhejiang Yuexiu University of Foreign Languages, the main research direction is higher education.