International Journal for Housing Science and Its Applications
Publish August 6, 2025. Volume 46, Issue 3 Pages 6210-6222

https://doi.org/10.70517/ijhsa463533

Research on Multi-dimensional Cluster Analysis and Intelligent Configuration Optimization for Educational Resources Big Data

Yanhui Liu[1,2,*]
[1] Academy of Innovation Education, Chongqing Open University, Chongqing, 400000, China
[2] Academy of Innovation Education, Chongqing Technology and Business Institute, Chongqing, 400000, China
Corresponding authors: (e-mail: liuyanhui@cqtbi.edu.cn).

Abstract In the era of big data, the unbalanced distribution of educational resources in colleges and universities leads to significant differences in the quality of education, and the traditional allocation method lacks scientificity and precision. In this study, the improved K-means clustering algorithm is used to conduct multi-dimensional analysis of students' learning behaviors and educational resources demand, and construct a resource allocation optimization model to achieve intelligent allocation of educational resources. Through clustering analysis of nine learning behavior indicators of 143 college students, the learners were classified into three groups: excellent learners, ordinary learners and risky learners, among which the excellent learners were outstanding in terms of video viewing time (168.43 minutes) and the number of times of chapter study (147.31 times). The correlation between PPT courseware and final test scores was found to be 0.62, indicating that courseware mastery had a significant effect on student performance. Based on this, an optimal allocation model of educational resources containing six indicators, including teacher-student ratio and number of subjects per teacher, was constructed and applied to the allocation of teacher and library resources in five universities. The results show that after allocating 2,000 teachers through the intelligent allocation model, the student-teacher ratio of each university changes from 12.09-19.43 before allocation to 12.09-13.08 after allocation, which realizes the equalization of resource allocation. The research results provide scientific basis for colleges and universities to teach students according to their abilities, and have important guiding value for promoting rational allocation of educational resources and improving teaching quality.

Index Terms Educational resources big data, K-means clustering algorithm, intelligent allocation optimization, learning behavior analysis, resource utilization efficiency, multi-dimensional clustering

I. Introduction

With the rapid development of technology, the field of education is experiencing a revolution driven by the Internet and big data technology [1]-[3]. Educational resources are no longer limited to the traditional textual form, but expanded to multimodal data including images, audio, video, etc [4], [5]. These rich multimodal educational resources bring new challenges for searching and utilizing them [6]. Therefore, studying the optimization of educational resource allocation based on multidimensional cluster analysis is not only of great practical significance, but also of vital importance in promoting the intelligent development of the educational field [7]. In the field of education, the diversity of resource data is increasing, covering a variety of modalities such as text, image, and video [8], [9]. These multimodal educational resources require us to be able to simultaneously process and integrate semantic information from different modal data. Through multimodal semantic fusion, the semantic information of data from different sources can be effectively integrated to construct a more comprehensive and accurate representation of educational resources [10]-[12]. This fusion can not only realize the joint representation of text, image, video and other data, but also present the teaching content more vividly, provide teachers and students with diversified teaching and learning methods, and meet the demand for personalized learning [13].

With the deepening of the integration of big data, the implementation of intelligent services of smart educational resources becomes more and more feasible [14]. In order to effectively promote the development of intelligent services for smart educational resources, it is necessary to systematically integrate resources such as data, technology and personnel, deeply explore the potential value of data, and commit to promoting the in-depth fusion between data and specific businesses, so as to give full play to the enabling role of data [15]-[17]. In addition, it is also necessary to strengthen the interaction and communication between intelligent educational resources and users, accurately meet the personalized needs of users, and thus significantly enhance the effectiveness of

intelligent educational resources services [18]-[20]. Through such a dual strategy, it can not only promote the continuous progress of intelligent educational resources intelligent services, but also provide users with more high-quality and efficient educational resources services [21].

As an important strategic resource of the country, the rationality of its allocation directly affects the quality and efficiency of education. At present, the problem of unbalanced distribution of educational resources among colleges and universities prevails, which is manifested in the obvious differences in faculty strength, teaching facilities, research conditions, etc., which restricts the improvement of the overall education level. Especially in the context of the big data era, the learning behavior data generated by students contains rich information, which not only reflects the learning status and effect of students, but also indirectly reflects the rationality of the allocation of educational resources. However, the traditional method of educational resource allocation is often based on experience or simple statistical analysis, which is difficult to accurately capture the differentiated needs of students, resulting in inefficient utilization of resources. Educational resource allocation should be student-centered and scientifically allocated to the characteristics and needs of different types of students, but due to the lack of in-depth analysis and understanding of students' learning behaviors, it is difficult to achieve accurate teaching. At the same time, the allocation of educational resources in colleges and universities involves multiple dimensions and objectives, how to achieve multi-objective optimization under the conditions of limited resources is also a complex issue. The “big” of educational resources big data does not mean the quantity is big, but emphasizes the “value” is big, that is, from the complicated educational resources data can find the mystery, diagnose the problem, and improve the effect of educational resources allocation. Through the cluster analysis of students' learning behavior, different types of learners and their characteristics can be identified, providing a scientific basis for the personalized allocation of educational resources. In addition, correlation analysis between educational resources can reveal the degree of influence of different resources on learning effects and guide the optimal allocation of resources. Currently, the research on educational big data analysis technology and optimization allocation model is still in the exploratory stage, and there is an urgent need for an intelligent allocation method that can comprehensively consider students' learning characteristics and multi-dimensional factors of educational resources.

Starting from educational resources big data, this study first adopts the improved K-means clustering algorithm to conduct multidimensional analysis of students' online learning behavior data to identify the characteristics of different types of learners and their educational resources needs. Through clustering analysis of nine quantitative indicators of three types of behaviors: content learning behavior, homework learning behavior and interactive learning behavior, we classify learners into different types and explore the characteristics of using educational resources and differences in the needs of each type of learner. Second, the correlation between different educational resources is analyzed to reveal the degree of their influence on learning effects. On this basis, an intelligent allocation optimization model of educational resources is constructed, taking into account multi-dimensional indicators such as teacher-student ratio, proportion of full-time teachers, number of subjects per teacher, value of fixed assets per student, per student expenditure on education and number of students, and adopting a multi-objective planning method to realize the optimal allocation of educational resources. Finally, the validity of the model is verified through examples to provide theoretical basis and practical guidance for the rational allocation of educational resources in colleges and universities.

II. Big Data Analysis of Educational Resources Based on K-Means Clustering

In this chapter, the improved K-Means clustering algorithm is used to study students' learning behaviors and their corresponding demand for educational resources, which provides a basis for the subsequent optimal allocation of educational resources.

II. A. Subjects of study

This study selected all the undergraduates enrolled in the history of gardening course offered in the fall semester of 2022-2023 at the College of Forestry, University of S. There were 143 undergraduates enrolled in the course, of which 135 were from the class of 2021, 5 were from the class of 2020, and 3 were from the class of 2019.

II. B. Classification of Online Learning Behavior and Data Sources

The categorization of online learning behaviors and the related factor indicators are directly related to teaching resources, teaching modes and instructional design and even the online education platform itself. However, at the same time, related studies also show that not all indicators have significant correlation or the same influence effect on learning effects. On the basis of previous studies, this study classifies online learning behaviors into three categories: content learning behaviors, assignment learning behaviors and interactive learning behaviors. And from the platform's "exam/homework", "results", "resources" and "analysis report" and other functional modules to

determine specific quantitative indicators and collect relevant data. In the "Exams/Assignments" module, a statistical report including user name, name, test score, test completion time, and test duration is exported from the record of each test. In the "Grade Results" function module, export statistical reports of data such as user name, name, total learning time, number of content resources used, and number of replies. In the "Resource Location" module, statistical reports of various content resource types, views, and content learning time are exported. Table 1 shows the statistical results of the teaching resources and related records released by the course to the platform.

Table 1: Statistical results of teaching resources and related records

Types of teaching resources	Acceptable range	Quantity/piece	Views/times	Study duration /min	Data source (Module)
Content resources	PPT courseware	35	5924	12437	Resource area
	Related content links	27	45	13	Resource area
	Excellent assignment PDF	14	1176	395	Resource area
Assignment resources	Chapter test (not included in the total score)	4	655(Answer)	3438	Exam/Assignment
	Chapter test	6	1215 (Answer)	12619	Exam/Assignment
	Final test	1	152 (Answer)	5214	Exam/Assignment
	Group assignment	3	Submit in groups		Exam/Assignment
	Independent operation	2	145 (Submit)		Exam/Assignment
Interactive resources	Questionnaire survey	1	124	124 (Reply)	Questionnaire survey
	Post	25	5622	1246 (Reply)	Academic results
	Blog	15	1815	0 (Reply)	Learning community /Blog

Ultimately, nine factors such as the number of content resources used, the length of video viewing, the number of chapter studies, the number of replies to posts, the length of reading, classroom activity, homework scores, classroom tests, and midterm tests were selected as quantitative indicators of learning behavior to be investigated in this study.

II. C.Research methodology

At present, the analysis tools for online learning behavior mainly include professional learning analysis tools, network analysis packages and visual network analysis software, or various data mining techniques relying on general software such as EXCEL, SPSS, etc., such as decision tree, regression analysis, social network analysis, artificial neural network, plain Bayes, support vector machine, time-series analysis, K-Means clustering and principal component analysis.

Online learning behavior is observable and measurable, therefore, this study starts from the two dimensions of observation of learning behavior and quantitative analysis of big data, on the basis of observation of three types of online learning behavior, namely, content learning behavior, homework learning behavior and interactive learning behavior, and in accordance with the above mentioned nine quantitative indexes of online learning behavior, after the data organization of EXCEL software, the data are imported into SPSS software to conduct data descriptive statistics, correlation statistics and cluster analysis, in order to further analyze and excavate the online learning behaviors of college students in the course of history of gardening, and to provide a basis for the subsequent course construction, teaching design and academic reform.

II. C. 1) K-means clustering algorithm

The K-means clustering algorithm [22] is aimed at dividing the data and this is achieved by dividing it into k clusters based on the input parameter k. The algorithm first randomly selects k points and uses them as the initial clustering centers, and then the difference from the sample to the clustering center is used as the basis, and classifies it under the category of the clustering center with the smallest difference until the new class obtained recalculates the clustering centers, ends the sample adjustment, and meets the condition that there is no change in the clustering centers of A and B. When the condition of the clustering center of A and B does not change, the clustering criterion function J_c reaches convergence.

The algorithm is a dynamic clustering algorithm that utilizes a batch-by-batch modification approach in the iterations, so that at any time an iteration occurs, the samples are checked and adjusted if an error occurs. Only

after the adjustment is completed can the clustering center be modified and the next round of iteration is started. If no error occurs, i.e., all samples are classified correctly, no adjustment is required and no modification of the clustering center is needed, the clustering criterion function J_c reaches convergence and the algorithm stops. The steps of the algorithm are as follows:

(1) Define n as a deterministic dataset and make $I=1$, select k initial clustering centers $Z_j(I), j=1,2,3,...,k$.

(2) Find the difference in distance from each sample to the cluster center $D(x_i, Z_j(I)), i=1,2,3,....,n, j=1,2,3,...,k$, if Eq. (1) is satisfied:

$$D(x_i, z_k(I)) = \min\{D(x_i, z_j(I))\} \tag{1}$$

Then $x_i \in w_k$.

(3) Find the error sum of squares criterion function J_c:

$$J_c(I) = \sum_{i=1}^{k}\sum_{k=1}^{n_j} \| x_k^j - Z_j(I) \|^2 \tag{2}$$

(4) Judgment: if $|J_c(I) - J_c(I-1)| < \xi$ then the algorithm stops. Instead $I = I+1$, solve for k new cluster centers, $Z_j(I) = \frac{1}{n}\sum_{i=1}^{n_j} x_i^j$ and repeat step (2).

II. C. 2) Optimizing the effectiveness of the K-means clustering algorithm

K-means clustering algorithm still has some problems when dealing with student stratification, such as too much emphasis on the initial value, so that the stratification will be affected by the data and lead to the existence of the defect of having the optimum, so this paper improves the standard K-means clustering algorithm to make it more effective.

(1) Data initialization denoising

K-means algorithm is easily affected by isolated points as well as noise, so this paper mainly focuses on this aspect when improving it. The distance sum between any point i and the rest of the points is calculated, and the distance sum is denoted as S_i, and then the distance average is calculated and denoted as H, in the case of $S_i > H$, this point is an isolated point. Also n denotes the dimension of the data in d and d denotes the sample data. I.e:

$$S_i = \sum_{j=1}^{n}\sqrt{\sum_{h=1}^{d}(x_{ih} - x_{jh})^2} \tag{3}$$

$$H = \sum_{i=1}^{n}\frac{S_i}{n} \tag{4}$$

The algorithmic procedure is as follows:

1) Scan the data set A once and find the distance mean sum of all data points and H distance sum S_i.

2) For any data point i, if the distance sum is greater than the distance average, i.e., $S_i > H$, i is an isolated point.

3) Remove all isolated points in A to obtain a new data set, denoted as A'.

Denote the mutual distance between any two points by $d = \sqrt{(x_2 - x_1)^2} + (y_2 - y_1)^2$, and use (x_1, y_1) and (x_2, y_2) to denote the coordinates of the two data points, where x denotes the horizontal coordinate and y denotes the vertical coordinate. The process of selecting the initial clustering centers can be divided into the following six steps:

1) First find the isolated points, find the distance of N objects from each other, and then output its distance matrix c_{id}, which is a matrix of $n \times n$.

2) Do $[t, c] = \max(\max(c_{id}))$ and compute the sum of the elements of each row of the matrix, i.e., the sum of the distances between all objects.

3) Do the operation $[q, 1] = \max(a)$ and compute the point with the largest sum of distances and note its position as 1, i.e. one of the isolated points.

4) Delete the isolated point obtained in the third step and repeat the first step of the operation to find the point whose sum of distances is the second largest in the output distance matrix of $N-1$ points, and this operation is repeated until the accuracy requirement is satisfied.

5) The M isolated points are deleted, and after the accuracy requirement is met, the remaining data points output their own distance matrix c_{id}.

6) Perform the $[t,c]=\max(\max(c_{id}))$ operation to compute the largest distance value and its column position, i.e., one of the two points with the largest distance, and perform the $[t,a]=\max(c_{id}))$ operation to compute the maximum value of the remaining points in the columns and record their row coordinates. Find the row coordinates of the data in column c, i.e., the remaining point of the two maximum points. The two points thus obtained are the initial cluster centers.

The initial centroids and center values of the clusters are entered and the operations are executed and the optimal clustering centers are obtained by a continuous iterative clustering algorithm. Determine whether the isolated points are to be excluded or not, and solve for the mutual distances between the isolated points as well as the selected clustering centers on the basis of the distance from the center most principle.

(2) Initial center of mass optimization

The clustering effect of the K-means algorithm is easily affected by factors such as the initial center of mass chosen and the initial data input, for which the approach of K cluster centers is envisioned in this paper.

Assuming that the number of data samples in the data set Q is n and using k to denote the number of clusters, the original K-means algorithm sets $S=\sum_{j=1}^{k}\sum_{C_j}d^2(x-m_j)$ as the objective function, and by which the computation of the minimum is performed. This function attaches importance only to intra-class distances, which are minimized while not considering inter-class distances, so the objective function can be changed to:

$$J(c,k)=\sqrt{W(c)W(c)+b(c)b(c)} \tag{5}$$

In the formula:

$$W(c)=\sum_{i=1}^{k}W(C_i)=\sum_{i=1}^{k}\sum_{x_i\in\delta_j}d^2(x,C_i) \tag{6}$$

$$b(c)=\sum_{1\le j\le i\le k}d^2(c_j,c_i) \tag{7}$$

C_i denotes the class, $W(c)$ denotes the sum of the intra-class distances of the k class, $d(x,C_i)$ denotes the in-class distance of the class C_i, $d(c_j,c_i)$ denotes the class distance of the class C_i and C_j class, c_i,c_j represent the class C_i,C_j, $b(c)$ is the sum of the interclass distances of the k class. If the objective function $J(c,k)$ is taken to the minimum, then the compactness within the class and the independence between classes will be enhanced.

Secondly, according to the defects of the initial center in the original K-means algorithm, the initial center point is determined according to the sample distribution.

Suppose there are n samples for Q, the number of variables is defined as p, the number of clusters is defined as k, and the data x,y are defined by the Euclidean distance [23]:

$$d(x,y)=\sqrt{(x_1-y_1)^2+(x_2-y_2)^2+...+(x_p-y_p)^2} \tag{8}$$

The distance between the sample x and the data set Q is denoted as:

$$d_{\min}(x,Q)=\min(d(x,y),y\in Q) \tag{9}$$

The maximum distance between the sample x and the data set Q is denoted as:

$$d_{\max}(x,Q)=\max(d(x,y),y\in Q) \tag{10}$$

Data samples are categorized into data sets with the restriction:

$$l=\frac{\max\limits_{1\le x,y\le n}(d(x,y))-\min\limits_{1\le i,j\le n}(d(i,j))}{k} \tag{11}$$

II. D.Analysis of the results of the study

The “big” of education resources big data does not mean the quantity is big, but emphasizes the “value” is big, that is to say, it can discover the mystery from the complicated education resources data, diagnose the problems, and

improve the effect of education resources allocation. Based on the research design, the results of the study are presented from three different perspectives: data characterization, cluster analysis and correlation analysis.

II. D. 1) Perspective 1: Characteristics of the school situation

Two methods of data characterization, the central tendency measure and the dispersion measure, were used to measure the basic academic information. Central tendency is a statistic that indicates the degree to which a set of data is converging towards a central value, and allows for the search for representative values in a set of data. The data dispersion metric, on the other hand, assesses the extent to which values are scattered and dispersed, giving a clear picture of the distribution of a set of data. In this case, a box-and-line diagram is used for visual presentation to provide key information about the location and dispersion of the data, including five statistics: minimum, first quartile, median, third quartile, and maximum, and the results of the specific analysis of the academic profile are shown in Figure 1.

Overall, first, the data centrality, the most stable in terms of the number of content resources used, the majority of students' content resources used in high numbers, indicating that the educational resources of the course can ensure that the majority of students' learning, homework scores, classroom quizzes and midterm tests of the three learning performance assessment data distribution is more centralized, reflecting that the level of learning within the classroom group tends to be consistent. Secondly, in terms of data discrete, all dimensions are distributed with the appearance of outliers, such as the content resources, there are serious behaviors of learning such as the use of a smaller number of content resources, which leads to a lower score for some students in this area. Video viewing time and reading time there is the phenomenon of some students' time is too high, there may be the phenomenon of students hanging time without watching. The low scores in homework scores, classroom quizzes and midterm tests, as well as the less-than-satisfactory assessment of students, may be due to the lack of learning initiative, self-discipline and self-management ability necessary for independent learning, which leads to the problems of inattention during the learning process, lack of active participation in teaching and learning interactions, and poor quality of homework, and triggers the phenomenon of polarization of learning results among different groups of students. Thirdly, the length of video viewing and the quality of homework. Third, the length of video watching and article reading, as the main task points of online learning, are still close to 0, which indicates that some learners in the course still have poor independent learning ability, low classroom participation, and cannot effectively complete the course requirements. Fourth, the number of replies and low classroom activity, online courses exist in the dilemma of less teacher-student interaction, students' participation is not extensive, and classroom activity is not active.

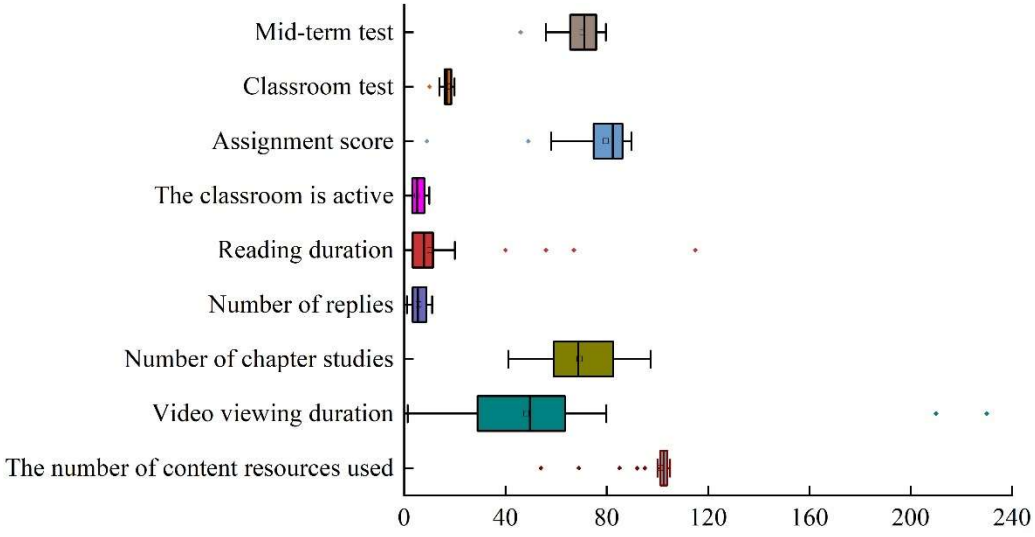

Figure 1: Student situation information

II. D. 2) Angle 2: Clustering of Learning Situations

The overall knowledge of the learner situation is achieved by describing the learner characteristic information, while learner clustering analysis can distinguish learners from within the group based on the different performance of their classroom behavior data. In this paper, the improved K-means algorithm is used for learner clustering analysis, and the value of contour coefficient k is shown in Figure 2. The whole folding line obviously accompanies the more clusters, the contour coefficient value is getting lower and gradually tends to the level, and when k=3, the clustering

effect is optimal. Combining the online learning behavior data and the results presented by clustering, the learner groups are divided into three categories: excellent learners, ordinary learners and risky learners, and the mean values of the behavioral characteristics of the learners within the clusters are organized as shown in Table 2.

The results show that the students in Cluster 3 perform very positively in all learning behaviors and have better data values than Cluster 1 and Cluster 2, and are classified as excellent learners. Students in Cluster 1 are not active in their learning behavior and do not pay attention to their learning and have lower data values than Cluster 2 and Cluster 3 and are classified as risky learners. Students within Cluster 2, on the other hand, had behaviors between Cluster 1 and Cluster 3, excelled in homework scores, classroom quizzes, and midterm tests, but were slightly deficient in task point completion and were classified as average learners.

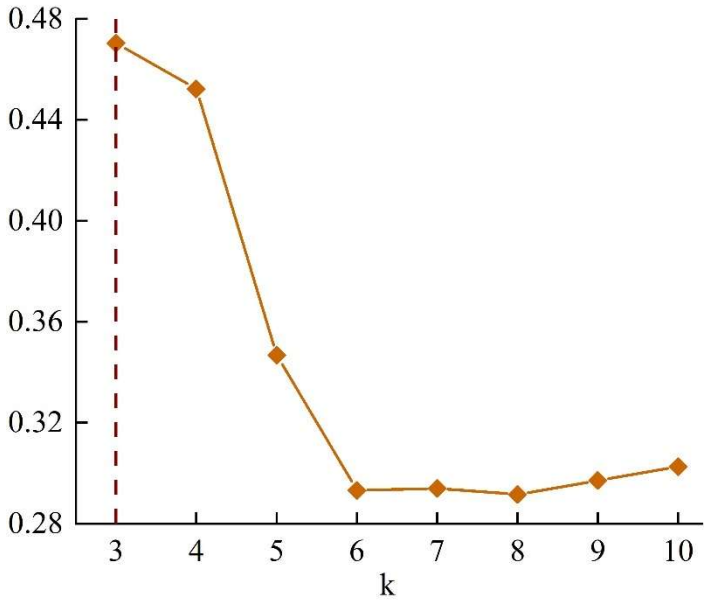

Figure 2: Contour coefficient: k value

Table 2: Statistics of the mean values of characteristics within clusters

Student behavioral performance	Cluster 1	Cluster 2	Cluster 3
The number of content resources used	95.32	97.94	100
Video viewing duration	11.54	92.31	168.43
Number of chapter studies	45.42	92.68	147.31
Number of replies	5.35	6.94	9.12
Reading duration	5.71	7.05	39.58
The classroom is active	3.08	5.29	0.65
Assignment score	78.04	82.61	82.86
Classroom test	14.65	16.73	15.28
Mid-term test	68.17	72.06	72.89

According to the clustering results, different learning behaviors contribute to different learning outcomes. In the group of excellent learners, students are highly engaged in the classroom and have excellent self-discipline, but still need to strengthen their internalization and construction of learning knowledge to improve their academic performance. Ordinary learners generally fulfill the teacher's goals for the classroom. At-risk learners, who are lagging behind in their learning, need to increase their self-discipline and be supervised and managed by the teacher. For the division of different learning groups, teachers can conduct scientific group teaching accordingly to avoid the chaos caused by random grouping. In group teaching, the grouping principle of “homogeneity between groups, heterogeneity within the group, complementary advantages” should be implemented as much as possible, contributing to the learning atmosphere in which the excellent learners lead the risky learners in the group. At the same time, the results of the cluster analysis also provide a basis for teachers to personalize the allocation of teaching resources according to the students' learning conditions.

II. D. 3) Perspective 3: Relevance of educational resources

In order to explore the key variables that influence the need for educational resources for courses that affect student learning, Pearson correlation coefficient [24] was used for educational resources correlation analysis to determine their relevance. The results of Pearson correlation coefficient are shown in Figure 3.

First, there is a weak correlation relationship between most of the factors in the course's educational resources. Second, it was found that the correlation between PPT coursework and final test scores was 0.62, which is a very high correlation, which reflects that students' knowledge of PPT coursework affects their final scores. Similarly, the correlation between group work and independent work was 0.54, and the correlation between chapter test 1, which is not included in the overall score, and chapter test 2, which is included in the overall score, was 0.59, which all showed some degree of positive correlation. According to this correlation result, strengthening the optimization of the combination and configuration of different educational resources within the online classroom provides a relevant basis for teachers to provide the optimal adjustment of instructional design, and at the same time provides certain reference for students to learn efficiently.

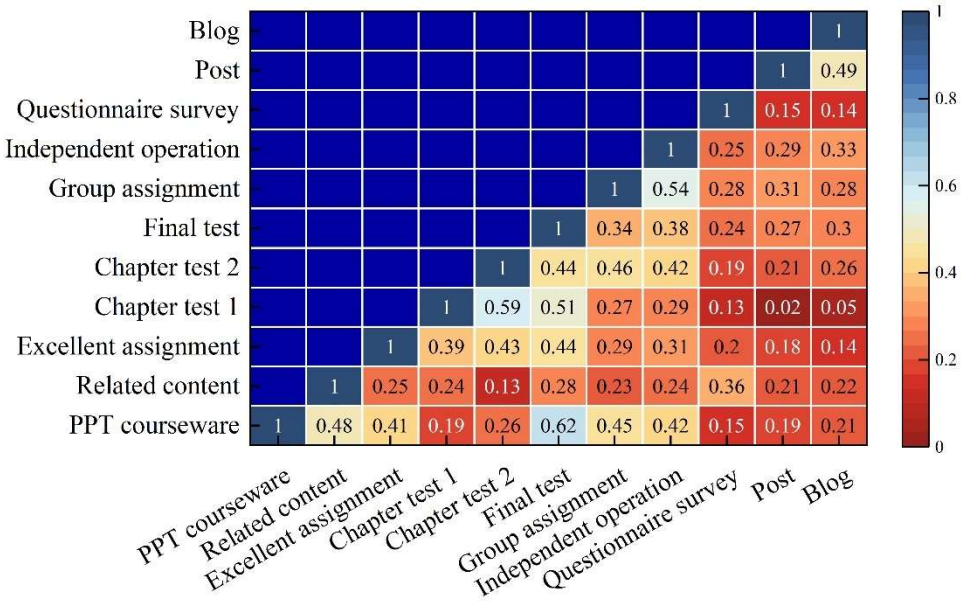

Figure 3: Pearson Correlation coefficient

III. Optimization of intelligent allocation of educational resources in colleges and universities based on academic data

In this chapter, on the basis of analyzing the demand of educational resources for students' learning by using the learning data, we further construct a resource allocation optimization model to realize the intelligent allocation optimization of educational resources.

III. A. Research methodology

III. A. 1) Selection of indicators for the optimal allocation of higher education resources

The problem of college education resource allocation involves more factors, and it can be classified as an optimization problem under multi-objective on a macro level.

The unbalanced allocation of university education resources is reflected in the input and output aspects, such as: regional per capita education input, the number of regional general colleges and universities, the average annual number of regional graduates and so on. In this paper, we measure the efficiency of the allocation of university education resources from three directions: per capita research output, per capita number of students trained by teachers, and per capita income from social services, and finally determine six indicators: XI teacher-student ratio (%), X2 the proportion of teachers among all faculty members, X3 the number of per capita number of subjects per teacher, X4 the value of per capita fixed assets at the end of the year, X5 the per capita expenditure on education, and X6 the number of enrolled students (10,000 people). In order to quantify the allocation efficiency, the values of the above indicators are used to calculate the resource utilization efficiency of universities, and the formula is:

$$S = C_1 * \underline{A_1} + C_2 * \underline{A_2} + C_3 * \underline{A_3} + C_4 * \underline{A_4} + C_5 * \underline{A_5} + C_6 * \underline{A_6} \tag{12}$$

where A_i is the score of A_i and C_i is the weight $(i=1,2,\cdots 6)$. Since the determination of weights is subject to individual subjective influence, this paper de-quantifies some indicators and establishes a multi-objective planning model.

Assume that $f(x)=(f_1(x),f_2(x)\cdots,f_9(x))$ denotes the objective function vector of the nine indicators of higher education resources, and $g(x)=(g_1(x)g_2(x)\cdots,g_9(x))$ denotes the vector of constraint functions for each indicator. The problem of optimal allocation of resources can be transformed into the following planning model:

$$\begin{gathered}\max=\{(Z_1=f_1(x),Z_2=f_2(x),\cdots,Z_9=f_9(x))\} \\ stg_i(x)\le 0,\quad i=1,2,\cdots,9 \\ x=(x_1,x_2,\cdots,x_9)\end{gathered} \tag{13}$$

In the feasible domain $Z=\left\{z\in R^n \mid z_1=f_1(x)z_2=f_2(x)\cdots,z_9=f_9(x)\right\}$, S^m represents the feasible domain of the decision space, and R^n denotes the feasible domain of the goal space.

III. A. 2) Establishment of a model for optimizing the allocation of educational resources

(1) Combination optimization of educational production factors in colleges and universities

Higher education resources are affected by multiple factors such as time, geography, social structure, etc., and its configuration problem is not a simple linear distribution. Among the factors of production, quality is the life and ultimate measure of higher education, and its ultimate goal is to maximize the benefits in terms of student output, knowledge output and social output.

Assuming that Z is the amount of educational output and $X_i(i=1,2,3,\cdots,n)$ is the different forms of the factors, then:

$$Z=f(x_1,x_2,\cdots,x_n) \tag{14}$$

$$P=\sum_{j=1}^{m}Z_j/\sum_{i=1}^{n}X_i \tag{15}$$

where P,X_i,Z_j denote the input-output ratio, the combination of different educational factors, and the amount of educational output, respectively.

As can be seen from the above equation, when P is larger, the combination of educational production factors is more reasonable, so the combination with the largest input-output ratio should be selected, but at the same time, with the increase of investment, the cost of education will also increase. Therefore, it is also necessary to consider the cost problem.

(2) The establishment of university education resources optimization model

The objective function is obtained through the above analysis:

$$xR_1=\sum_{i=1}^{n}x_j/\sum_{i=1}^{n}y_i\ \max R_2=\sum_{i=1}^{n}y_i/\sum_{i=1}^{n}z_i\ \max R_i=\sum_{i=1}^{n}D_i/\sum_{i=1}^{n}y_i \tag{16}$$

$$xR_s=\sum_{i=1}^{n}E_j/\sum_{i=1}^{n}y_j\ \max R_s=\sum_{i=1}^{n}F_j/\sum_{i=1}^{n}x_j\ \max R_6=\sum_{i=1}^{n}x_j \tag{17}$$

where, $X_{wv},Y_{wv},D_{wv},E_{wv},F_{wv},H_{wv}\ge 0$, and $\forall uv$ at the same time, the following constraints should be satisfied:

$$\begin{gathered}S.t.\quad \alpha_{11}\le\sum_{i=1}^{nj}x_{ij}\le\alpha_{12}\alpha_{21}\le\sum_{i=1}^{nj}y_{ij}\le\alpha_{22}\alpha_{31}\le\sum_{i=1}^{nj}y_{ij}\le\alpha_{32} \\ \alpha_{41}\le\sum_{i=1}^{nj}y_{ij}\le\alpha_{42}\alpha_{51}\le\sum_{i=1}^{nj}y_{ij}\le\alpha_{52}\end{gathered} \tag{18}$$

The above model is built for each indicator of higher education, R_1 is the school's student-teacher ratio, R_2 is the proportion of full-time faculty to all staff, R_3 is the total number of subjects per teacher ($R\ \&\ D$), R_4 is the total value of per pupil fixed assets at the end of the year, R_5 is the per pupil total expenditure on education, and R_6 is the total number of students enrolled in school. Where n_j denotes the number of colleges and universities in the j region, and the other indicators for the j region are represented by the annual totals $\sum_{i=1}^{i}U_i$.

III. A. 3) Solving the model

(1) Adaptation function selection

Maximizing the use of resources in order to achieve maximum efficiency is the essence of the optimal allocation of higher education resources. In order to express the change trend of the objective function more clearly, this paper selects the inverse of the objective function as the objective function, and uses the fast non-inferiority sorting method to get the non-inferiority set, and the main idea is as follows: based on the concept of Pareto solution, comparing each solution in the population with other solutions to get a Pareto solution set, which is used to find the first Pareto frontier solution set, and is recorded as F1. At this time, we put aside the first Pareto frontier and continue to compare according to the concept of Pareto solution to get the second Pareto frontier F2. Frontier, continue to follow the Pareto solution concept for comparison, to get the second Pareto front F2, repeat the above process, until all the Pareto front $\{F_1, F_2, \cdots\}$.

(2) Algorithm Design

The algorithm design flow is shown in Fig. 4.

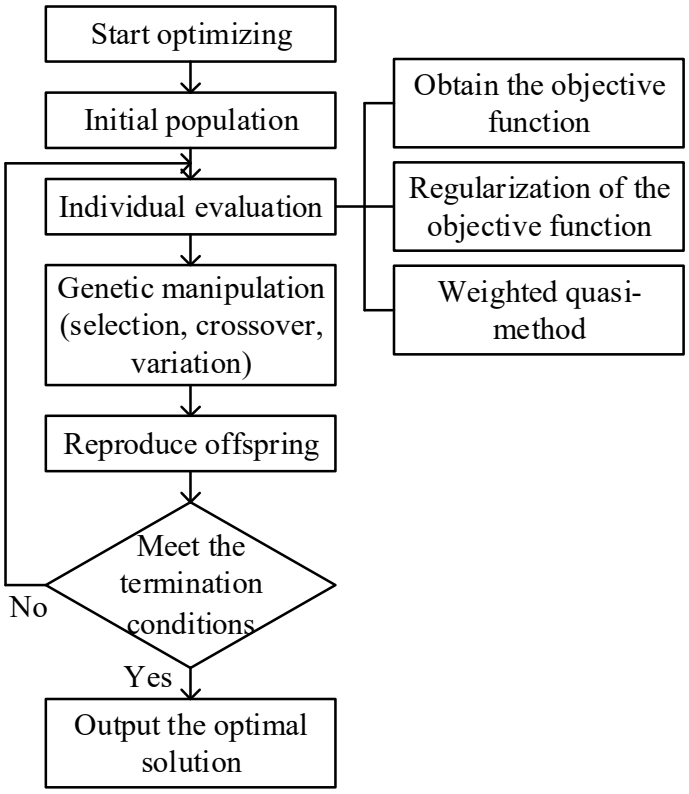

Figure 4: Algorithm design process

III. B. Research process and analysis of results

This section takes five colleges and universities as the object to carry out the study of optimization of educational resource allocation, assuming that there are currently 2,000 teachers and 120,000 new books need to be allocated to these five colleges and universities, in order to improve the allocation of educational resources in these five colleges and universities in the statistics, and to achieve a more balanced state.

First of all, the assumption of 2000 teachers for the allocation of educational resources, the five colleges and universities of the student-teacher ratio in descending order and draw a bar chart as shown in Figure 5, reflecting the five colleges and universities of the student-teacher ratio of the existence of large gaps.

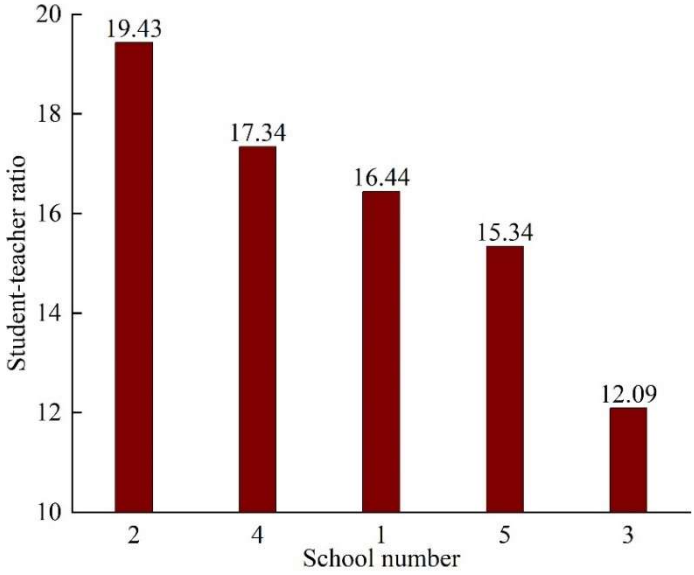

Figure 5: The student-teacher ratio situation before the allocation

The optimal allocation of the 2000 teachers to be allocated is carried out by using the optimization model of educational resource allocation, and the purpose of the allocation model is to equalize the status quo of the teachers' resources in each university so that they can achieve balanced allocation, and the results of the allocation are shown in Table 3.

It can be seen that four colleges and universities have been allocated new teachers to different degrees, and only one college and university has not been allocated new teachers in the early period when the student-teacher ratio is relatively small, and the student-teacher ratio of these five colleges and universities has been equalized through intelligent allocation. From the comparison of the student-teacher ratio before and after the allocation of each university, it can be obtained that through the hypothetical resource allocation of this intelligent allocation model, it can be more effective to improve the problem of unbalanced distribution of educational resources in terms of teachers' resources in each university, so that after the allocation of teachers' resources have reached a more even situation, and improved the problem of unbalanced educational resources in terms of teachers.

Table 3: The changes before and after the allocation of teachers in each university

Colleges and universities	Number of students	The number of teachers	Student-teacher ratio	Assigned number of teachers	The student-to-teacher ratio after allocation
1	29600	1800	16.44	527	12.72
2	27200	1400	19.43	718	12.84
3	17900	1480	12.09	0	12.09
4	21500	1240	17.34	450	12.72
5	27000	1760	15.34	305	13.08

The optimal allocation of the 2000 teachers to be allocated is carried out by using the optimization model of educational resource allocation, and the purpose of the allocation model is to equalize the status quo of the teachers' resources in each university so that they can achieve balanced allocation, and the results of the allocation are shown in Table 4.

It can be seen that four colleges and universities have been allocated new teachers to different degrees, and only one college and university has not been allocated new teachers in the early period when the student-teacher ratio is relatively small, and the student-teacher ratio of these five colleges and universities has been equalized through intelligent allocation. From the comparison of the student-teacher ratio before and after the allocation of each university, it can be obtained that through the hypothetical resource allocation of this intelligent allocation model, it can be more effective to improve the problem of unbalanced distribution of educational resources in terms of teachers' resources in each university, so that after the allocation of teachers' resources have reached a more even situation, and improved the problem of unbalanced educational resources in terms of teachers.

Table 4: The changes before and after the distribution of books in each university

Colleges and universities	Number of students	Number of books/ten thousand	The average number of books per student	The number of books distributed	The average number of books per student after allocation
1	29600	261.3	88.28	36158	89.50
2	27200	249.0	91.55	29482	92.63
3	17900	173.0	96.65	3754	96.86
4	21500	211.8	98.53	1736	98.59
5	27000	227.6	84.30	48870	86.11

The changes in the number of books and the per capita book ratio before and after the book allocation in each university are plotted as a line comparison graph as shown in Figure 6.

It can be seen that the average number of books per student in these five colleges and universities is not very balanced after the allocation. The main reason for this situation may be that the total number of books in these five colleges and universities before the allocation of the base number of books are several million, the number is relatively large, so the number of books assumed in the model for the allocation of books is still too small compared to the number of books owned by the colleges and universities before the distribution of the 120,000 new books to significantly improve the current situation of the imbalance in the number of books per pupil is more difficult. This shows that in order to fundamentally improve the imbalance of educational resources due to the large base, it is necessary to anticipate whether the distribution of educational resources will be equalized or not, instead of remedying the problem after the fact.

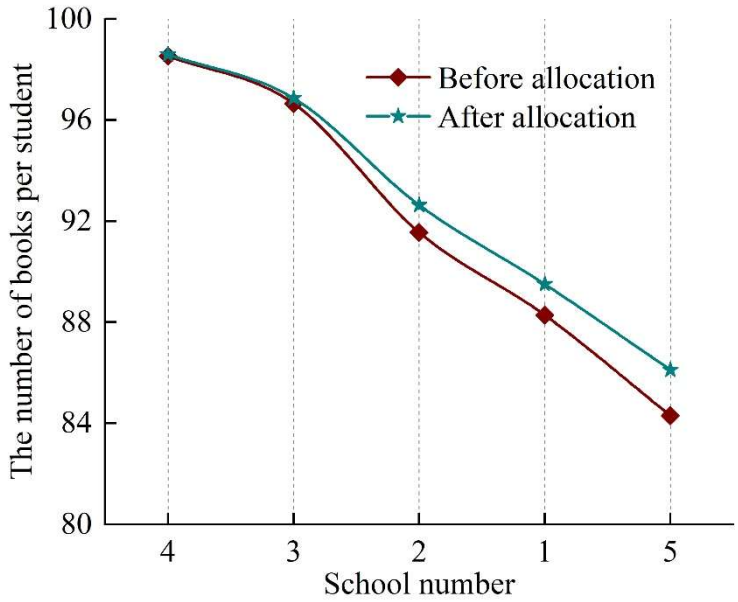

Figure 6: Comparison of the average number of books per student in each university

IV. Conclusion

The study confirmed the effectiveness of big data analytics on educational resources in identifying students' needs and optimizing resource allocation through multidimensional cluster analysis and intelligent allocation optimization research.K-means cluster analysis divided learners into three groups, in which at-risk learners in terms of the number of content resources used (95.32), the number of times chapters were studied (45.42), and the midterm test (68.17 points) performance lagged behind, suggesting that such students need increased study supervision and individualized instruction. Pearson's correlation analysis reveals that there are many kinds of correlations between course resources, especially the correlation coefficient between group work and independent work reaches 0.54, and the correlation coefficient between chapter tests not included in the total grade and chapter tests included in the total grade is 0.59, which provides a scientific basis for the allocation of the combination of educational resources.

The intelligent allocation optimization model of educational resources shows good results in practical application. After the allocation of 120,000 books to five universities, the gap between the average number of books per student

is narrowed from 14.23 books before allocation to 12.48 books, and the resource allocation tends to be balanced. The model breaks through the limitations of traditional resource allocation methods and realizes data-driven accurate allocation.

In the future, educational resource allocation should pay more attention to the mining and application of learning data, and carry out differentiated resource allocation for the characteristics of different learning groups. Colleges and universities should establish a regular learning behavior data collection and analysis mechanism to continuously optimize the allocation of educational resources and promote the improvement of educational quality. At the same time, they should deepen the research on multi-objective optimization model to further improve the science and precision of resource allocation.

Funding

This work was supported by the Science and Technology Research Program of Chongqing Municipal Education Commission (Grant No. KJQN202304003).

References

[1] Zheng, S. Y., Jiang, S. P., Yue, X. G., Pu, R., & Li, B. Q. (2019). Application research of an innovative online education model in big data environment. International Journal of Emerging Technologies in Learning, 14(8).

[2] Dahdouh, K., Dakkak, A., Oughdir, L., & Ibriz, A. (2020). Improving online education using big data technologies. In The role of technology in education. IntechOpen.

[3] Huda, M., Anshari, M., Almunawar, M. N., Shahrill, M., Tan, A., Jaidin, J. H., & Masri, M. (2016). Innovative teaching in higher education: The big data approach. Tojet, 1210-1216.

[4] Olivier, J. (2019). Short instructional videos as multimodal open educational resources in a language classroom. Journal of Educational Multimedia and Hypermedia, 28(4), 381-409.

[5] Linder, R., & Falk-Ross, F. (2024). Multimodal resources and approaches for teaching young adolescents: a review of the literature. Education Sciences, 14(9), 1010.

[6] Girón-García, C., & Fortanet-Gómez, I. (2023). Science dissemination videos as multimodal supporting resources for ESP teaching in higher education. English for Specific Purposes, 70, 164-176.

[7] Azevedo, B., Pereira, M., & Araújo, S. (2022, November). Designing a multilingual, multimodal and collaborative platform of resources for higher education. In International Conference on ArtsIT, Interactivity and Game Creation (pp. 391-404). Cham: Springer Nature Switzerland.

[8] Lin, A. (2012). Multilingual and multimodal resources in genre-based pedagogical approaches to L2 English content classrooms. English–A changing medium for education, 79, 103.

[9] Ting, K. Y. (2014). Multimodal Resources to Facilitate Language Learning for Students with Special Needs. International Education Studies, 7(8), 85-93.

[10] Tsoukala, C. K. (2021). STEM integrated education and multimodal educational material. Advances in Mobile Learning Educational Research, 1(2), 96-113.

[11] Govender, R., & Rajkoomar, M. (2021). A multimodal model for learning, teaching and assessment in higher education. COVID-19: Interdisciplinary explorations of impacts on higher education, 57.

[12] Song, J., Chen, H., Li, C., & Xie, K. (2023). MIFM: Multimodal Information Fusion Model for Educational Exercises. Electronics, 12(18), 3909.

[13] Chango, W., Lara, J. A., Cerezo, R., & Romero, C. (2022). A review on data fusion in multimodal learning analytics and educational data mining. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 12(4), e1458.

[14] Chen, N. S., Yin, C., Isaias, P., & Psotka, J. (2020). Educational big data: extracting meaning from data for smart education. Interactive Learning Environments, 28(2), 142-147.

[15] Han, J. (2023). Rational Planning of Educational Resources Based on Big Data Fusion. International Journal of Web-Based Learning and Teaching Technologies (IJWLTT), 18(2), 1-12.

[16] Ma, H., Wang, X., & Liang, D. (2015). Research on analysis of multi-source data fusion model in Smart classroom. Journal of Information Technology and Application in Education, 4, 82.

[17] Gong, T., & Wang, J. (2023). A data-driven smart evaluation framework for teaching effect based on fuzzy comprehensive analysis. IEEE Access, 11, 23355-23365.

[18] Wu, S., Cao, Y., Cui, J., Li, R., Qian, H., Jiang, B., & Zhang, W. (2024). A comprehensive exploration of personalized learning in smart education: From student modeling to personalized recommendations. arXiv preprint arXiv:2402.01666.

[19] Peng, H., Ma, S., & Spector, J. M. (2019). Personalized adaptive learning: an emerging pedagogical approach enabled by a smart learning environment. Smart Learning Environments, 6(1), 1-14.

[20] Yang, Q. (2019). Research on Active Service Model with Personalized Education Resources. Asian Journal of Contemporary Education, 3(1), 95-104.

[21] Embarak, O. H. (2022). Internet of Behaviour (IoB)-based AI models for personalized smart education systems. Procedia Computer Science, 203, 103-110.

[22] Yu* Lin, Bai Yujie & Shankar Achyut. (2024). Design of network security monitoring system based on K-means clustering algorithm. Intelligent Decision Technologies,18(4),3105-3118.

[23] Rustam Mussabayev. (2024). Optimizing Euclidean Distance Computation. Mathematics,12(23),3787-3787.

[24] Spyros Tserkis, Syed M. Assad, Ping Koy Lam & Prineha Narang. (2025). Quantifying total correlations in quantum systems through the Pearson correlation coefficient. Physics Letters A,543,130432-130432.