J. COMBIN. MATH. COMBIN. COMPUT. 127a (2025) 4095--4109

Journal of Combinatorial Mathematics and Combinatorial Computing
www.combinatorialpress.com/jcmcc

Combinatorial Press

Research on the Optimization Path of English Classroom Interaction in Colleges and Universities Based on Big Data

Siyuan Yuan[1,✉]

[1] *School of Foreign Languages, Henan University of Animal Husbandry and Economy, Zhengzhou, Henan, 450000, China*

ABSTRACT
This project focuses on the classroom interaction of college English and proposes a framework for optimizing college English classroom interaction by integrating big data. Taking the behavioral analysis layer as the entry point, using PSO's improved K-mean clustering algorithm, we focus on analyzing the specific application of data mining technology in students' learning behavior. Then we conduct experiments on two classes of students in a university, design classroom behavioral coding to analyze classroom interaction behavior, and explore the application effect of this English classroom interaction optimization pathway. The students were divided into six categories through cluster analysis, with focused learners (22%) and continuous learners (36%) having the highest fidelity scores and the largest proportion, and the analysis of students' learning behaviors can provide a reference for teachers' classroom teaching. The composition of the English interaction optimization classroom changes from teacher-led to student-led in the traditional classroom, the teacher-student speech curves intersect each other and both appear four peaks, showing good classroom scope and teacher-student interaction effect, and the path of interaction optimization in the English classroom based on big data is practicable.

Keywords: big data, PSO, K-mean clustering, learning behavior analysis, interaction optimization

1. Introduction

In today's digital era, big data technology is penetrating into all fields at an unprecedented speed, and education is no exception. As an important part of cultivating high-quality internationalized talents, English classrooms in colleges and universities are facing new opportunities.

Big data technology brings opportunities for English classes in colleges and universities in four main ways. First, it can realize personalized teaching. Through the analysis of a large amount of students' learning data, teachers can accurately understand each student's learning characteristics, strengths and weaknesses, so as to formulate personalized learning programs for them and provide

✉ Corresponding author.
E-mail address: helenysy@126.com (S. Yuan).
Received 01 March 2024; Revised 25 May 2024; Accepted 10 November 2024; Published Online 15 April 2025.
DOI: 10.61091/jcmcc127a-231

targeted tutoring and support [1-2]. Second, optimize the allocation of teaching resources. Big data can help schools and teachers to understand students' needs and use of different teaching resources, so as to allocate teaching resources more rationally, such as teaching materials, courseware, online courses, etc., and improve the efficiency of resource utilization [3-5]. Thirdly, it promotes the innovation of teaching methods. Based on the analysis results of big data, teachers can try new teaching methods and means, such as flipped classroom, blended teaching, etc., to better meet students' learning needs [6-7]. Fourth, to improve the scientificity and accuracy of teaching assessment. Big data can comprehensively and dynamically record students' learning process and performance, providing a richer and more objective basis for teaching assessment and avoiding the one-sidedness and subjectivity of traditional assessment methods [8-9]. Therefore, many colleges and universities have established digital teaching platforms to collect students' learning data, such as learning hours of online courses, completion of assignments, and test scores. These data provide some reference for teachers to understand the learning status of students, but there are still some problems in the practical application.

First, the collection and analysis of data are not comprehensive and in-depth [10]. Most platforms only focus on students' learning outcome data, and pay less attention to behavioral data in the learning process, such as learning paths and the choice of learning strategies. Secondly, the application of data is mainly focused on teaching assessment, and has limited effect on optimizing and improving teaching content and methods [11-12]. Furthermore, teachers' and students' knowledge and application of big data technology needs to be improved, and many teachers do not know how to effectively use data to make teaching decisions, and students are not clear about how to adjust their learning strategies based on data analysis [13-15]. For this reason, there is a need to optimize the interaction between big data and the English classroom in colleges and universities.

Based on big data technology, the study proposes an English classroom interaction optimization framework, which contains three layers: the data collection layer, the behavior analysis layer and the teaching application layer. Focusing on the teaching discussion of student learning behavior analysis in the behavior analysis layer, taking the English classroom learning data of the university learning platform as the research object, after optimizing the K-mean algorithm using the particle swarm algorithm, it performs the cluster analysis of the collected data and calculates the loyalty of the students' learning behaviors, and divides the students into different loyalty level categories. Subsequently, two parallel classes of students in a university were selected for teaching practice, and after formulating the English classroom interaction behavior coding, the recovered classroom levels were analyzed using Nvivo 11.0 software to obtain the classroom behavior ratio and teacher-student speech curves of the traditional English classroom and the interaction-optimized classroom, and to test the teaching effect of the speech classroom interaction optimization framework through the comparative analysis of teacher-student interactions of the two modes of classrooms.

2. A framework for optimizing English classroom interactions based on big data

The effect of classroom interaction plays a crucial role in teaching evaluation, and the improvement of education and teaching quality has been the main goal of teaching reform. With the development of information technology, teaching interactive platforms and other digital devices are applied to education and teaching, prompting changes in teaching methods, learning tools and learning styles.

At present, the teacher-student interaction in college English classroom exists problems such as single interaction subject, low interaction participation, irrational use of interaction function and failure to maximize the interaction time, etc. Therefore, this paper proposes the optimization path of college English classroom interaction based on big data technology. The English classroom interaction optimization framework based on big data is shown in Figure 1. The framework is divided into three layers, the data collection layer, the behavioral analysis layer and the teaching application layer, and the three layers are connected to each other for cyclic optimization.

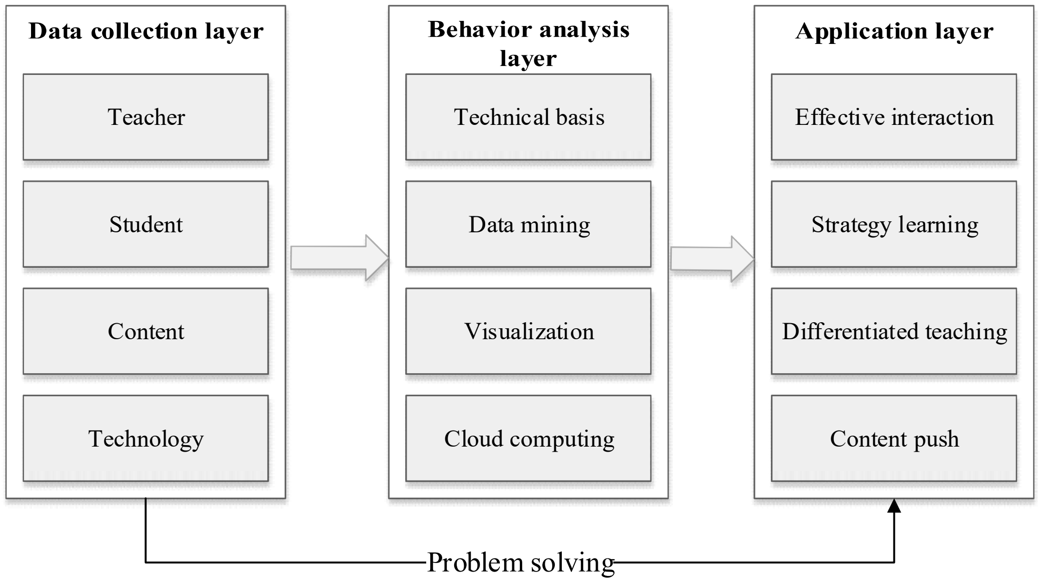


Fig. 1. English classroom interaction optimization framework based on large number

2.1. Data collection layer

The basic support for the optimization of teacher-student interaction, the main basis for the learning situation and student behavior, the data comes from the main body of classroom interaction, that is, the teacher, the student, the technology and the content of the teaching, the model in the digital classroom environment, for the collection of data to provide an interactive teaching support platform, the mobile terminal APP, manual entry and other information systems. Through the current situation of the construction of the smart classroom environment, it can be seen that the platform system of the smart classroom is still through the integration of many manufacturers' systems and equipment, the degree of integration is not high, and the problem of learning data detection and interaction has not really been solved, i.e., the means of data collection is still the same as that in the digital classroom. The optimization of teacher-student interaction needs to be judged by integrating the interaction data and learning results in the learning process to provide data support for interaction interventions.

Interaction data generated during the learning process is divided into three categories of data: emotional, expressive and operational data, of which likes, tags, favorites, etc. are emotional data. The expression category includes text, images and comments. The operation category contains information such as interface switching and dwell time.

Learning outcome data include a test completed by students, learning tasks, improved learning works, participation in polls, and mutual assessment among students and teacher evaluation.

2.2. Behavioral analysis layer

Behavior analysis layer is mainly based on learning analysis technology as the core, supported by the combination of theory and technology, processing and analyzing the collected data, and landing the interaction optimization and intervention strategy in the actual classroom teacher-student interaction.

Technical support includes data mining, visualization, cloud computing and other technologies, among which, data mining is used to process and discover information from process data such as students' button frequency and viewing interface. Visualization technology is used to present the results of data analysis to teachers visually so that they can adjust the interaction method. Cloud technology, on the other hand, is used to store data, share data information, and provide real-time data resources.

The application of learning and analysis technology is divided into "data source", "data processing", "data analysis" and "result presentation", and the data is processed and cleaned and transformed and then returned to "data analysis", and the results are output to the teaching application layer after taking appropriate analysis methods.

2.3. Teaching application layer

Teaching application layer mainly adjusts the status quo of classroom interaction in time, enriches the form of interaction in the smart classroom, and reaches effective teaching. Through the interaction intervention, the effective interaction, strategic learning and differentiated teaching are achieved. Effective interaction is the feedback and diagnosis of teacher-student interaction, and the precise and effective adjustment of the interaction form and timing of interaction to enhance the learning effect. Strategic learning is based on analyzing students' learning process data to push resources for students and adjust learning steps to make teaching effective. Differentiated teaching is to effectively locate students' learning progress and provide teachers with effective analysis of learning conditions.

3. Analysis of online learning behavior based on data mining

Teacher-student interaction behaviors in the English classroom in higher education require classroom data collection, analysis, pedagogical application, and problem solving. The main purpose of the behavioral analysis layer is to conduct in-depth analysis and mining of data in order to provide more refined decision-making. In the behavioral analysis layer, a range of techniques and tools, such as data mining, machine learning, and artificial intelligence, are usually used to process and analyze the data in order to discover the patterns and laws behind the data. In this chapter, data mining techniques are used to analyze the historical data of learning behaviors of English language students in higher education through clustering in order to identify different types of learning behaviors and analyze them.

3.1. Cluster analysis methods

3.1.1. K-mean clustering algorithm. K-mean algorithm is a classical clustering algorithm, the algorithm is first set by the algorithm writer or user to artificially set the number of clusters K, the final result of clustering is to determine the location of the center of mass of the K clusters and form K clusters, the algorithm so that the similarity measure between the points (data) within each cluster is minimized, and the similarity measure between clusters is as large as possible.

K-mean clustering algorithm in the distance measure can choose a variety of forms, such as Euclidean distance, Manhattan distance, Minkowski distance, etc., the use of different distance measures, only in the calculation of distance in that step of the calculation method is different, the rest of the algorithm steps are the same. Which distance method is used in practical applications should be selected according to the characteristics of the actual application.

The metric used in the K-mean clustering algorithm to optimize the location of the center of mass is the intra-cluster error variance (SSE). For a specified k clusters, the smaller the distance between the points in the cluster, the corresponding SSE is correspondingly smaller and the clustering is more effective.

Let the data set be $A=\{a_1,a_2,\ldots a_i\}$, where $a_1,a_2\ldots a_i$ denotes each of the i data records in the data set. The initial center of mass is randomly generated for the first time, denoted by $K_1=\{K_1,K2_1,\cdots Kj_1\}$, where Kj denotes the j th center of mass and Kj_1 denotes the point where

Kj is located for the first time. After the center of mass is generated, traverse from a_1, a_2 to a_i to calculate their respective distances to $K1_1, K2_1, \ldots Kj_1$, and judge a_i which distance from Kj_1 is the closest to divide a_i into the cluster centered on that Kj_1. Calculate the SSE_1 of the first time, as in equation (1), which is the sum of the squares of the distances a_i from the center of mass Kj_1 that one is at:

$$SSE_1 = \sum_{m=1}^{j} \sum_{a_{m1} \in Km_1} \left\| a_{m1} - Km_1 \right\|^2 \tag{1}$$

Where SSE_1 denotes the first SSE, m denotes the variable representing the number of centers of mass at the time of programming, which takes integer values from 1 to j, a_{m1} denotes all the data belonging to the m th center of mass after the first center of mass is determined, Km_1 denotes the m th center of mass determined for the first time, and $a_{m1} \in Km_1$ denotes that a_{m1} is classified into clusters centered on the center of Km_1.

The mean value of the points within each cluster is calculated and it is taken as the new center of mass K_2:

$$K_2 = \left\{ K_2, K_2, \ldots Kj_2 \right\} \tag{2}$$

$$Kj_2 = mean\left(a_{m1} \right) \tag{3}$$

The $mean(\ )$ in Eq. (3) represents the mean value function, by which the center of mass K_2 of the second time can be determined, and the data in A are reassigned to the new cluster according to the principle of closest distance to calculate the SSE_2 of the second time:

$$SSE_2 = \sum_{m=1}^{j} \sum_{a_{m2} \in Km_2} \left\| a_{m2} - Km_2 \right\|^2 \tag{4}$$

Calculate the gap ΔSSE_2 between SSE_2 and SSE_1:

$$\Delta SSE_2 = \left| SSE_1 - SSE_2 \right| \tag{5}$$

To determine whether ΔSSE_2 meets the error requirement, set δ as a set positive number, which is the threshold value for error judgment and is determined according to the actual situation. If ΔSSE_2 is less than δ, i.e:

$$\Delta SSE_2 < \delta \tag{6}$$

Then the error of clustering is considered to meet the accuracy requirement and the algorithm stops. If it does not meet the requirements, then calculate the mean value of the points in each cluster with K_2 as the center of mass to form a new center of mass K_3, according to the principle of closest distance to the data in A and then assigned to the new cluster did not calculate the third time SSE_3, calculate ΔSSE_3 and determine whether to terminate or loop again until the end of the algorithm.

3.1.2. Particle Swarm Optimization Algorithm. Particle Swarm Optimization (PSO) algorithm is one of the more classical intelligent optimization algorithms for groups of intelligent optimization algorithms. The basic idea is that when a flock of birds is searching for food in a new place, at the beginning, each bird moves along the place where it thinks there may be food and notes down the place where it sees the most food in the search process, and then when the flock converges later and shares the information it has found, it will know the location where the most food exists at the moment, and then each bird will then re-conduct the search for food based on the information it searches for and the information shared by the team. Then each bird will search for food again based on the information it has searched for and the information shared by the team, and after several times of information sharing, the flock will be able to find the location where the most food exists.

The optimization method of particle swarm optimization algorithm mainly imitates the feeding process of bird population, firstly, the particle swarm optimization algorithm equates the problem to be optimized into a flock of birds that are feeding, then the solution space of the problem to be optimized is equated to the flight space of the birds in the process of feeding, and finally, each particle in the solution space of the problem to be optimized by the particle swarm optimization algorithm will be regarded as each bird in the flight space of the whole flock of birds. In the end, each particle in the solution space of the problem to be optimized by the particle swarm optimization algorithm will be regarded as each bird in the flight space of the whole flock of birds, that is, each potential solution of the problem to be optimized.

In the particle swarm optimization algorithm, since each potential solution to the problem can be replaced to a certain extent by using a particle that is equivalent, and the fitness value of each equivalent replaced particle can be computed by using the fitness function, which computes the fitness value of the corresponding particle:

Assuming that there is a D-dimensional retrieval space and a population X consisting of n particles, the position $X_i = \left(X_{i1}, X_{i2}, \ldots, X_{iD}\right)$ of the ith particle in the retrieval space can be used to represent a potential solution to the optimization problem, and the respective fitness value of each particle's position X_i can be computed based on the objective function. Continuing to set the velocity of the ith particle as $V_i = \left(V_{i1}, V_{i2}, \ldots, V_{iD}\right)$, the individual extreme value of the particle as $P_i = \left(P_{i1}, P_{i2}, \ldots, P_{iD}\right)$, and the population extreme value of the whole population as $P_g = \left(P_{g1}, P_{g2}, \ldots, P_{gD}\right)$, it should be noted that when using the particle swarm optimization algorithm, the distance and direction of the particle to move in the population are determined by the velocity of the particle itself, and the particle's velocity and positional superiority are adjusted because of each iteration of the search for optimization, so the particle's positional updating formula and velocity updating formula are:

$$V_{id}^{k+1} = \omega V_{id}^{k} + c_1 r_1 \left(P_{id}^{k} - X_{id}^{k}\right) + c_2 r_2 \left(P_{gd}^{k} - X_{gd}^{k}\right) \tag{7}$$

$$X_{id}^{k+1} = X_{id}^{k} + V_{id}^{k+1} \tag{8}$$

Among them, Eq. (7) is the particle's velocity update formula, and Eq. (8) is the particle's position update formula. In Eq. (8), ω represents the inertia weight, $d = 1, 2, \ldots, D$, $i = 1, 2, \ldots, n$, k represent the current number of iterations of the whole particle swarm algorithm, c_1 represents the individual learning factor of particles, c_2 represents the group learning factor of particles, both of which can also be collectively called acceleration constants, and r_1 and r_2 are random numbers randomly distributed in the region of $[0,1]$, which are used to increase the randomness of retrieval.

The particle's velocity update formula (7) contains a total of three parts: the inertial part, the cognitive part, and the social part. Among them, the inertial part consists of the inertial weight ω and the particle's own displacement V_{id}^{k}, which indicates the particle's dependence on its own current motion state. The cognitive part consists of the individual learning factor c_1, random number r_1 and the position difference between the current position P_{id}^{k} and the optimal position X_{id}^{k} that it has ever reached, which indicates the degree of the particle's thinking about itself and its reliance on the results of its own exploration. The social part consists of the group learning factor c_2, random number r_2 and the position difference between the current position and the optimal position that the group has ever reached, which indicates the particle's thinking about the group's result and the degree of trust on the group's exploration result.

Meanwhile, in order to avoid the particle exceeding the search range during the search, so the position of the particle will be limited in the interval $\left[-X_{\max}, X_{\max}\right]$ and the speed will be limited in the interval $\left[-V_{\max}, V_{\max}\right]$.

3.1.3. Improved K-mean clustering algorithm. The traditional K-mean clustering algorithm is weak in finding the global optimal solution, while the particle swarm algorithm is similar to the bird foraging, according to the law of iterative repetition, the comparison of the most value, narrowing the region until the optimal solution is found, compared to an intelligent algorithm with better global search ability, and both have the characteristics of simplicity and fast convergence. In order to adapt to the demand for higher accuracy of clustering results in the new environment, we can try to combine particle swarm algorithm with K-mean clustering algorithm for clustering analysis. Therefore, this paper introduces the particle swarm algorithm to improve the K-mean cluster clustering algorithm on the basis of the K-mean cluster clustering algorithm.

The process of improving K-mean clustering algorithm is as follows:

1) Firstly, an initialization of the parameters is carried out, because it is necessary to assign the initial position and velocity to the particles.

2) Calculate the individual optimal position of the particle swarm and the group optimal position fitness value.

3) Update the velocity and position of the particles.

4) Calculate the individual optimal position and population optimal position fitness values of the particle swarm again and compare them.

5) Optimize the traditional K-mean clustering algorithm. For each new particle generated, optimization is performed according to the K-mean clustering algorithm. The clustering centers of the particles are numbered and sorted, the Euclidean distance is calculated to minimize the spacing of particles in the cluster, and the attribution to the particles and the category in which they should be classified are determined.

6) If the end condition is reached, output the result. Otherwise, recalculate the clustering center and update the fitness of the particles, and turn to step (2) to continue iterating until the nearest neighbor law is met and the final clustering is achieved.

3.2. Learning Behavior Classification Indicators

In RFM customer classification analysis method, customers are distinguished by three behavioral indicators, namely, proximity R, frequency F, and value M. This method is mainly applicable to traditional sales scenarios, and relevant indicators need to be redesigned for learner classification in online learning platforms. By analyzing the behavioral data of learners in online learning, the RFL indicator system for online learner classification is designed. Where R is the number of days between the learner's last learning point in the e-learning platform and the current time, known as learning proximity. F is the sum of learners' behavioral records of learning in the e-learning platform, which include learning behaviors such as interactively answering questions, watching course videos, leaving messages and exchanging information, and taking e-notebooks, and is referred to as the frequency of learning.L is the cumulative number of minutes of learning behaviors that a learner performs through the e-learning platform, and is referred to as the length of learning.

By calculating the weight of each indicator in each category and following the given formula, the learners' loyalty score to the English course is calculated. This helps teachers and online learning platform administrators to adopt appropriate teaching strategies and teaching modes in time to improve the stability of students' English learning. The formula for calculating the loyalty of online learning behavior is shown in formula (9):

$$RFL_{(loyalty)} = -\alpha\overline{R} + \beta\overline{F} + \gamma\overline{L} \tag{9}$$

where α,β,γ is the weight of the categorical indicator R,F,L. $\overline{R},\overline{F},\overline{L}$ is the R,F,L mean value.

3.3. Analysis of Students' Learning Behavior

3.3.1. Data processing. The English course in the online learning platform of a university was selected as the specific research object. The course was offered during the period from September 1, 2023 to December 30, 2023, and the learning data corresponding to the indicators of learning proximity (R), learning frequency (F), and learning duration (L) of 100 learners were counted respectively.

3.3.2. Cluster analysis. For the standardized data, the PSO improved K-mean algorithm is used to cluster the data, and the number of clusters ranges from 2 to 10, and the scores of the number of clusters are calculated by using the variance ratio criterion method, and the results of score visualization are shown in Fig. 2, with the horizontal coordinate being the number of clusters, and the vertical coordinate being the corresponding score of each kind of number. The highest score is 870 when the number of clusters is 6, i.e., it is most appropriate to divide the data of students' online learning behaviors into 6 classes.

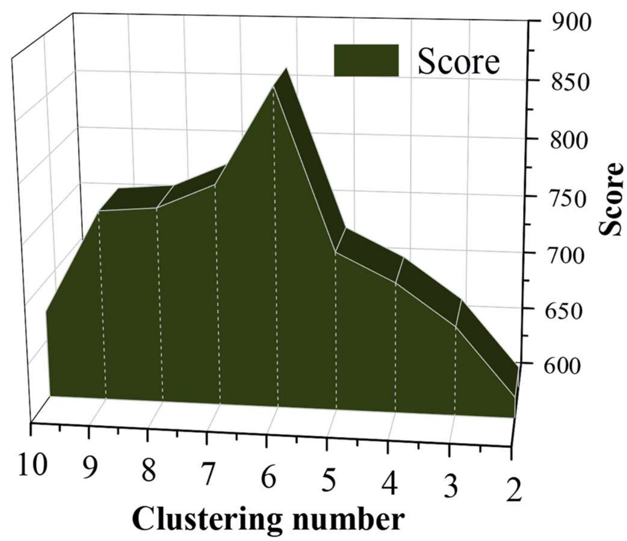


Fig. 2. Visual results of scoring

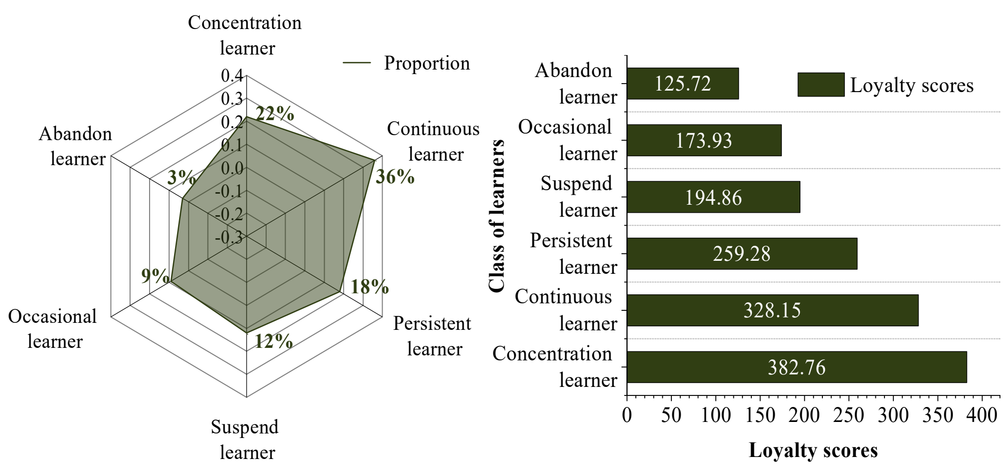
The PSO modified K-means algorithm was used to classify the student online learning behavior data into six categories and calculate the mean score for each category. The learning categories generated from the cluster analysis are shown in Table 1. The overall proximity mean, frequency mean and duration mean for students in English courses were 15.19 days, 44.96 times and 374.72 points, respectively.

Table 1. Learning class generated by cluster analysis

Class	Number	Approximate mean/day	Frequency mean/times	Time/minutes
1	22	2.11	125.36	951.13
2	36	6.96	116.99	939.72
3	18	7.89	72.18	679.29
4	12	41.34	60.07	544.47
5	9	14.13	53.94	422.58
6	3	79.46	21.09	209.96
Total	100	15.19	44.96	374.72

Loyalty was calculated for each category and based on the loyalty scores, Fig. 3(a) and Fig. 3(b) show the proportion of the distribution of the number of people in each cluster as well as the loyalty scores for each category, respectively.The six categories are focused learners, continuous learners, persistent learners, suspended learners, occasional learners, and abandoned learners. Focused learners and persistent learners had the largest percentages, 22% and 36%, and both had loyalty scores of 320 or more. Occasional learners and abandonment learners have the lowest percentage, both less than 10%.

The results of the cluster analysis and the learners' loyalty to the course can help teachers and platform administrators develop reasonable teaching strategies to improve the interactive effect and classroom quality of English classroom teaching. For example, focused learners have a high learning duration and frequency, and their latest online learning time is quite short from the current time. They are usually very focused on learning and will continue to invest a lot of time and energy in the learning process. In order to keep them focused and achieve better learning outcomes, teachers can provide them with more in-depth English learning materials, challenging learning tasks and more feedback to stimulate their interest and motivation.



(a)Proportion (b)Loyalty scores

Fig. 3. The proportion and loyalty scores of all kinds of different learners

4. Effectiveness of the application of the interaction optimization framework in the English classroom

In this chapter, based on the research on the optimization of English classroom interactions in higher education based on big data, the proposed framework is applied to English teaching, and the effectiveness of its application is discussed.

4.1. Subjects of study

In order to understand the characteristics of students' behaviors, teachers' behaviors, and teacher-student interactions in traditional English teaching and interaction-optimized classrooms, two parallel classes with comparable levels and numbers of students in a certain university were selected, and interaction-optimized practice was conducted in the English classroom for the students of one of them to obtain the video data of classroom teaching through the learning platform. At the same time, we obtain the classroom teaching video of the other class that adopts traditional teaching, and the time of each classroom session is about 40~50 minutes.

4.2. Research tools

4.2.1. Behavioral observation tools. The ITAS system was slightly modified in the light of actual classroom teaching to develop a coding list of interactive behaviors in the English classroom, including four categories: teacher verbal behaviors, student verbal behaviors, no verbal or technological, and technological behaviors. Among the specific behaviors:

Teacher verbal behaviors: lecturing X1, instructing X2, guiding and instructing X3, encouraging and praising X4, criticizing X5, evaluating and giving feedback X6, adopting opinions X7, asking open-ended questions X8, and asking closed-ended questions X9.

Student verbal behavior: passive response X10, active response X11, discussion X12, sharing telling X13.

No speech or technology: thinking about the problem X14, support X15, teaching chaos X16.

Technological behavior: teacher operates technology X17, student operates technology X18, technology acts on student X19.

4.2.2. Behavioral analysis tools:

1) Video analysis software Nvivo11.0. Nvivo11.0 software is a very powerful qualitative research and analysis software, which can handle documents, videos, audios, pictures and other kinds of data, and it is used in this study to complete the coding and analysis of English classroom videos.

2) Characteristic Curve Plot. The curve graph takes time as the horizontal coordinate and the proportion of teachers' and students' speech acts in each minute as the vertical coordinate, and draws the curve graph of speech act characteristics to describe the change of the proportion of teachers' or students' speech acts with time.

4.3. Findings

4.3.1. Classroom Interaction Node Coverage. The Nvivo11.0 software was used to code the resulting English classroom videos according to the English classroom behavior coding system, and then the Nvivo11.0 charting tool was used to draw graphs to get the coverage of each classroom-specific behavior and interaction type.

The results of the comparative analysis of the node coverage of interaction behaviors in the English classroom are shown in Figure 4. In the English traditional classroom, the behavior with the highest ratio of total interaction behaviors is teacher-alone behavior (68.3%), followed by teacher-student interaction (21.5%), and then teacher-technology interaction (5.5%), no interaction (0.19%), student-student interaction (0.15%), student-alone behavior (0.8%), student-technology interaction (0.5%).

The behaviors that accounted for the highest ratio of total interaction behaviors in the English interaction-optimized classroom were student-student interactions (40.10%), followed by teacher-student interactions (34.7%), and then student-technology interactions (11.8%), teacher-alone behaviors (6.7%), no-interaction behaviors (4.2%), teacher-technology interactions (1.9%), and student-alone behaviors (0.6%).

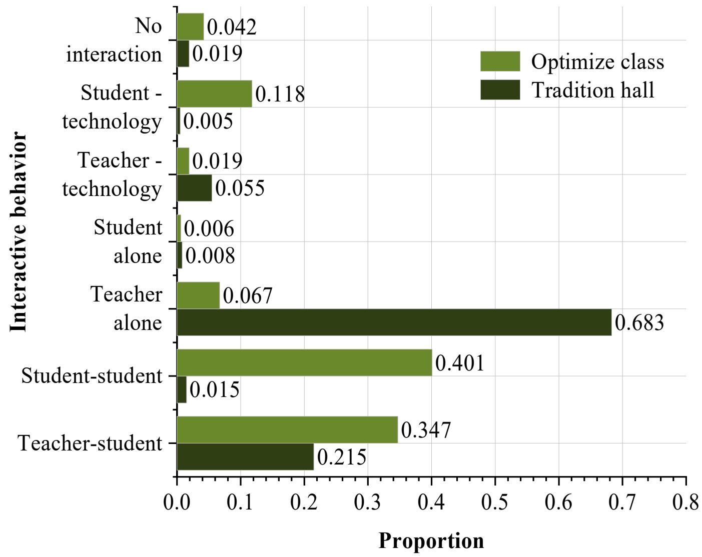


Fig. 4. The comparison results of the interaction behavior node coverage in English class

4.3.2. Comparison of Classroom Behavior Rates. The ratios of teachers' verbal behavior, students' verbal behavior, classroom silence (no verbal or technical behavior), and technical behavior in each classroom were measured, and the data were imported into SPSS for analysis, and the comparative analysis of the ratios of English classroom behaviors is shown in Table 2.The results found that there is a difference in the ratios of classroom behaviors between the traditional classroom of English and the interaction-optimized classroom (P<0.001), which suggests that classrooms with different instructional models of the English language curriculum have different ratios of teacher-student classroom behaviors. The composition ratio of teacher and student classroom behaviors is different, with the highest ratio of teacher verbal behaviors in traditional classroom being 85.81% and the highest ratio of student verbal behaviors in interaction-optimized classroom being 50.68%.

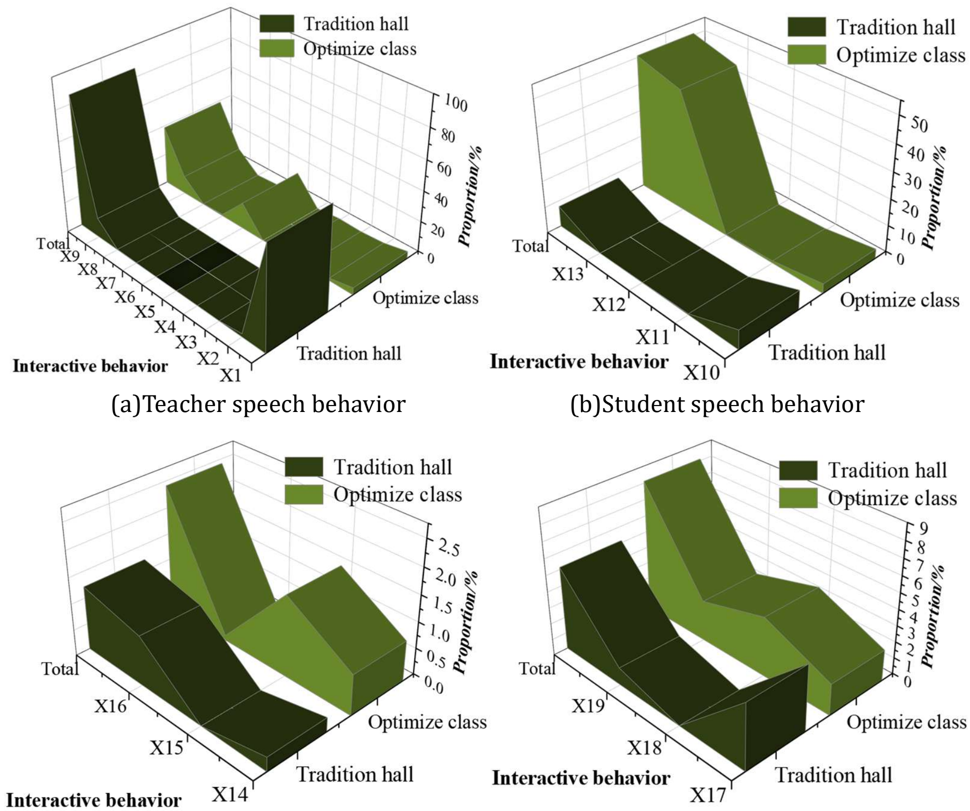
Table 2. Comparative analysis of English classroom behavior ratio

Item	Tradition hall		Optimize class		χ^2	P
	Frequency	Ratio/%	Frequency	Ratio/%		
Teacher speech behavior	635	85.81%	402	38.80%	537.14	<0.001
Student speech behavior	58	7.84%	525	50.68%		
Classroom silence	9	1.22%	25	2.41%		
Technical behavior	38	5.14%	84	8.11%		

The ratios of specific interaction behaviors in the English classroom are shown in Figure 5, (a) to (d) represent the ratios of teacher speech behavior, student speech behavior, classroom silence, and technology behavior, respectively. Specifically analyzing the verbal behavior in the classroom, most of the teachers' verbal behavior in the traditional English classroom is lecturing on the classroom content, accounting for 68.11%, and although there are more questions, accounting for 12.84%, they are mainly closed questions. In terms of encouragement and praise, teachers seldom encouraged and praised students, indicating that teachers in traditional classrooms were deficient in creating a positive interactive atmosphere.

In the interaction-optimized classroom, teachers' speech acts are no longer mainly lecturing, but evaluating feedback on students' sharing and telling behaviors in the classroom (17.18%), and teachers' feedback on students' learning can efficiently and purposefully improve students' knowledge and skills. In terms of encouragement and praise, teachers encouraged and praised students many times, which increased students' motivation and initiative to a certain extent and created a good interactive environment for English classroom teaching. Most of the students' behaviors were in sharing to tell their group results, which accounted for 45.95%, and this way not only practiced the students' ability to tell on the stage, but also improved the students' motivation to learn.

In terms of classroom silence and technology behavior, the traditional classroom is mainly the teacher operating the technology (4.05%), indicating that in the traditional classroom the teacher shows the lecture content in a single way. In contrast, in the interactive optimization classroom, the percentages of students manipulating technology, technology acting on students, and teachers manipulating technology were 3.67%, 2.51%, and 1.93%, respectively, indicating a better integration of technology and classroom in this classroom.



(a)Teacher speech behavior (b)Student speech behavior

(c)Classroom silence (d)Technical behavior

Fig. 5. The ratio of specific interactions in English classes

4.3.3. Classroom teacher-student speech curves. The curves of the teacher-student speech act ratio in the traditional English classroom and the interaction-optimized classroom are shown in Figures 6 and 7, respectively. In the traditional English classroom, the teacher's per-minute verbal behavior rate exceeds 60%, and the number of times reaching 90% or more is 14, indicating that the teacher controls most of the classroom time, and his/her teaching pace is relatively tight. In the students' verbal behavior rate, the students' per-minute speech rate is low, accounting for 0% to 40%, and the classroom interaction is ineffective.

In the English Interaction Optimization Classroom, the graphs of the teacher's speech behavior and the students' speech behavior rate form two relatively symmetrical sawtooth curves, which shows that the teacher and students jointly dominate the classroom in this classroom and the time allocation is relatively even. In this classroom, there were four peaks in both student and teacher verbal behavior. In this model, students actively participated in classroom activities, teacher-student interaction and student-student interaction were active, and the classroom learning atmosphere was favorable, which promoted students' deep learning.

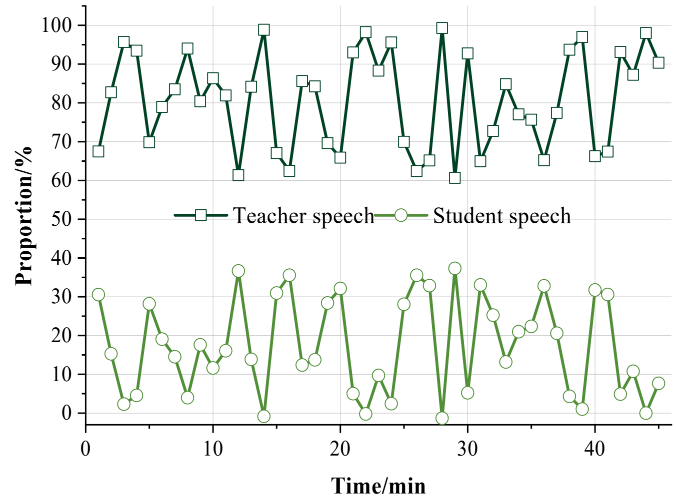


Fig. 6. The teacher-student speech curve of traditional English class

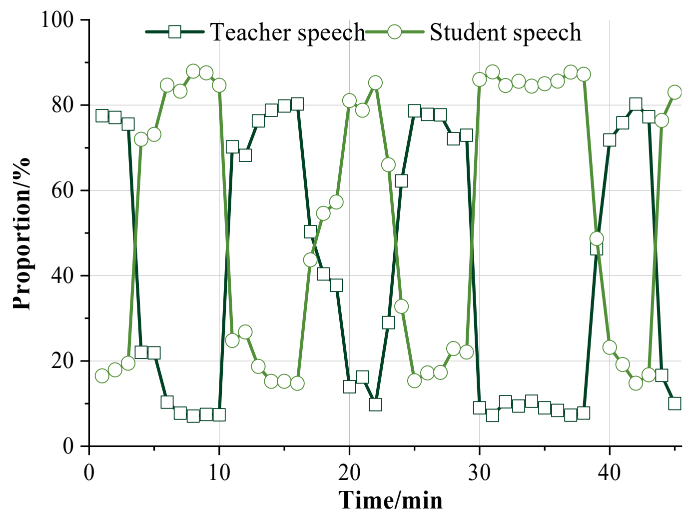


Fig. 7. The teacher-student speech curve of reciprocal optimization class

5. Conclusion

The embodiment of effective classroom teaching is the effectiveness of teaching and learning, and teacher-student interaction is one of the forms of classroom teaching, which is a key link to improve the quality of education and teaching. This paper combines big data technology to propose a framework for optimizing English classroom interaction in colleges and universities. Data mining technology is used to analyze students' online learning behaviors by examples, and the simulation application of the English interaction optimization framework is carried out to explore its application effect.

The PSO improved K-mean algorithm is used to cluster the online learning behaviors of students in English classrooms, and six different loyalty categories of students are obtained, which are focused learners, continuous learners, persistent learners, paused learners, occasional learners, and abandonment learners, and the loyalty scores of the first two categories of students are the highest, which are all over 320, and they account for the largest number of students in the sample, which are respectively 22% and 36%. Teachers can refer to the students' loyalty scores and use different teaching interactions for different categories of students to enhance their English learning.

There is a significant difference in classroom behaviors between traditional English classrooms and interaction-optimized classrooms at the 1% level, and the English classrooms applying the interaction-optimized framework have better teacher-student interaction effects. The traditional English classroom and interaction-optimized classroom have the highest composition of teacher speech acts and student speech acts respectively, with corresponding percentages of 85.18% and 50.68%. The traditional speech classroom teacher is mainly the teacher's individual lecturing behavior, with the teacher's speech accounting for more than 60%, while the teacher-student speech curve in the interaction-optimized English classroom shows a jagged shape, indicating that the classroom scope under the approach is good and the teacher-student interaction is strong, and it can promote the development of quality classroom interaction. The effectiveness of the application of the English classroom interaction optimization framework is verified.

References

[1] Regan, P. M., & Jesse, J. (2019). Ethical challenges of edtech, big data and personalized learning: twenty-first century student sorting and tracking. Ethics and Information Technology, 21(3), 167-179.

[2] Tapalova, O., & Zhiyenbayeva, N. (2022). Artificial Intelligence in Education: AIEd for Personalised Learning Pathways. Electronic Journal of e-Learning, 20(5).

[3] Williamson, B. (2017). Big Data in Education: The digital future of learning, policy and practice. SAGE Publications.

[4] Ray, S., & Saeed, M. (2018). Applications of Educational Data Mining and Learning Analytics Tools in Handling. Applications of Big Data Analytics: Trends, Issues, and Challenges, 135.

[5] Meng, W., Yu, L., & Zhu, Y. (2024). Quality Improvement Model of English Teaching in Universities Based on Big Data Mining. Journal of Electrical Systems, 20(3s), 506-518.

[6] Zhong, J., Li, Z., Hu, X., Wang, L., & Chen, Y. (2022). Effectiveness comparison between blended learning of histology practical in flipped physical classrooms and flipped virtual classrooms for MBBS students. BMC Medical Education, 22, 1.

[7] Crawford, R., & Jenkins, L. E. (2018). Making Pedagogy Tangible: Developing Skills and Knowledge Using a Team Teaching and Blended Learning Approach. Australian Journal of Teacher Education, 43(1).

[8] Yunita, A., Santoso, H. B., & Hasibuan, Z. A. (2021, June). Research Review on Big Data Usage for Learning Analytics and Educational Data Mining: A Way Forward to Develop an Intelligent Automation System. In Journal of Physics: Conference Series (Vol. 1898, No. 1, p. 012044). IOP Publishing Ltd..

[9] Costa, E. B., Fonseca, B., Santana, M. A., de Araújo, F. F., & Rego, J. (2017). Evaluating the effectiveness of educational data mining techniques for early prediction of students' academic failure in introductory programming courses. Computers in Human Behavior, 100(73), 247-256.

[10] Al Rawashdeh, A. Z., Mohammed, E. Y., Al Arab, A. R., Alara, M., & Al-Rawashdeh, B. (2021). Advantages and Disadvantages of Using e-Learning in University Education: Analyzing Students' Perspectives. Electronic Journal of e-Learning, 19(3).

[11] Morgulev, E., Azar, O. H., & Lidor, R. (2018). Sports analytics and the big-data era. International Journal of Data Science and Analytics, 5(4), 213-222.

[12] Meng-yue, C., Dan, L., & Jun, W. (2020). A Study of College English Culture Intelligence-Aided Teaching System and Teaching Pattern. English Language Teaching, 13(3).

[13] Ramadhona, N., Putri, A. A., & Wuisan, D. S. S. (2022). Students' opinions of the use of quipper school as an online learning platform for teaching english. International Transactions on Education Technology (ITEE).

[14] Sledgianowski, D., Gomaa, M., & Tan, C. (2017). Toward integration of Big Data, technology and information systems competencies into the accounting curriculum. Journal of Accounting Education, 100(38), 81-93.

[15] Hancock, J. T., & Khoshgoftaar, T. M. (2020). CatBoost for big data: an interdisciplinary review. Journal of big Data, 7(1), 94-94.