International Journal of Multiphysics
Volume 18, No. 2, 2024
ISSN: 1750-9548

Exploration of Information Practice of Constructed English Learning Platform Based on K-means Algorithm and Improved Apriori Algorithm

Wang Yu
Jilin Animation Institute, Changchun, Jilin 130000, China

Abstract

This paper explores the application of the constructed English learning platform based on k-means algorithm and improved Apriori algorithm in information teaching. Through the loss function optimization, the learning rate adjustment, the precise selection of the initial center point, and the improvement of the Apriori algorithm, this study realizes the accurate analysis of the learner characteristics and the personalized recommendation of the learning path. The experimental results show that the application of these techniques significantly improves the accuracy and efficiency of learners' clustering, the accuracy and user satisfaction of dynamic learning path recommendation, and the timeliness and personalization of interactive learning feedback. In addition, generic metrics such as improved user retention and overall satisfaction further demonstrate the success of the platform in enhancing user engagement and satisfaction. This study provides an effective technical path for the construction of personalized learning environment, and provides an important theoretical and practical basis for further optimizing the function of learning platform, improving user learning experience and promoting personalized learning practice.

Keywords: Constructed English learning platform, k-means algorithm optimization, improved Apriori algorithm, personalized learning path recommendation, learners group.

1. Introduction

In the information practice of constructed English learning environment, data mining technology, especially the application of k-means algorithm and Apriori algorithm, shows its core role in personalized teaching strategy. Historical studies, such as the improvement of the k-means algorithm, mainly focus on optimizing the clustering accuracy and processing efficiency, and adapting to different dataset properties by dynamically adjusting the algorithm parameters. However, the research of Apriori algorithm focuses on the efficient mining of correlation rules, such as improving the operation speed by reducing the generation of candidate sets and optimizing the storage structure of the algorithm. Although these technologies have been widely used in the business and social sciences, the application of educational technologies, especially constructive English learning platforms, has not fully demonstrated their potential [1].

This study aims to deeply explore the integrated application of k-means and improved algorithm of Apriori in a constructive English learning platform to achieve fine-grained analysis of learner characteristics and personalized recommendation of learning pathways. By optimizing the loss function and learning rate of k-means algorithm, this study improves the stability and accuracy of learner data clustering; meanwhile, the improved Apriori algorithm enhances the accuracy and efficiency of learning content correlation analysis by optimizing the support and confidence calculation. The integration of these technologies not only improves the teaching adaptability and personalized service ability of the platform, but also significantly increases the participation and satisfaction of learners through experimental verification. Therefore, this study not only provides an effective technical realization path for the constructed English learning platform, but also provides a new theoretical and application framework for the future educational informatization practice, which is of great significance for promoting the development of personalized learning environment [2].

2. Research Foundation

2.1 K-means algorithm

2.1.1 K-means loss function optimization of the clustering algorithm

Among the many algorithms of data mining technology, k-means clustering algorithm is widely used in learner data analysis because of its high efficiency and simplicity, and especially plays a vital role in the process of constructing personalized learning environment. The k-means algorithm enables clustering of the data by iteratively finding the center points of the k clusters in the dataset and assigning each data point to the nearest cluster center. The key to the algorithm is to minimize the sum of the distances from each point to the center of its cluster, i. e., the optimization loss function

$$J=\sum_{i=1}^{k}\sum_{x\in S_i}\left\|x-\mu_i\right\|^2 \tag{1}$$

S_i Where represents the set of all points in the i th cluster, is the central point of the cluster. μ_i S_i

However, the performance of standard k-means algorithms is limited by the local optimal solution of the loss function when handling complex and variable learner data, resulting in insufficient accuracy and stability of the clustering results. For this problem, the optimization of the loss function becomes a key way to improve the algorithm efficiency and clustering quality. By introducing regularization terms or adopting more complex distance measures, the algorithm can effectively avoid falling into the local optimal solution, and then improve the accuracy and generalization ability of clustering [3]. In addition, considering the diversity of learner characteristics and dynamic changes, dynamically adjusting the weight parameters in the loss function to enable the algorithm to adapt to the changes of learner behavior, which is important for realizing personalized learning path recommendation in a real sense.

2.1.2 K-means clustering algorithm learning rate optimization

Learning rate optimization of K-means clustering algorithm is a key factor to improve the convergence speed and accuracy of the algorithm, adopting a gradient descent based optimization strategy and introducing adaptive learning rate adjustment methods such as the update rules used in the Adam optimizer. This method considers not only the first moment estimation (i. e., the mean of the gradient), but also the second moment estimation (i. e., the uncentralized variance of the gradient), thus achieving a more refined and efficient learning rate adjustment [4].

Considering that the K-means algorithm itself does not use the gradient descent directly, but for the sake of discussion, we can analogy the iterative updating process of K-means to the parameter updating process in gradient descent and borrow the idea of adaptive learning rate. Specifically, the update of the cluster center in each iteration can be seen as optimizing the following objective functions:

$$J=\sum_{i=1}^{k}\sum_{x\in S_i}\left\|x-\mu_i\right\|^2 \tag{2}$$

Based on this, an adaptive learning rate adjustment strategy like that in Adam optimizer is introduced, where the update rule of cluster centers can be redesigned as:

$$\mu_i^{(t+1)}=\mu_i^{(t)}-\frac{\alpha}{\sqrt{\hat{v}_t}+\grave{o}}m_t \tag{3}$$

m_t $\hat{v}_t$ Where, and represent the exponential moving average of all gradients up to the current iteration, respectively, is a small constant avoiding dividing zero. In this way, the learning rate is adaptively adjusted to update the cluster center with a more optimized step size in each iteration, thereby accelerating the convergence process and improving the clustering accuracy. Although gradient descent is not directly applied in the native K-means algorithm, it can provide a new perspective and method for learning rate optimization of K-means algorithm by comparing its updating process to gradient descent and borrowing from the optimization strategy of adaptive learning rate. This advanced optimization method is expected to improve the algorithm performance when processing large-scale and high-dimensional learner data, providing more precise support for the recommendation of personalized learning paths [5].

2.1.3 K-means clustering algorithm

In the traditional method, the random selection of initial center points is easy to lead to the instability of clustering results, especially in the face of high dimensional data, the clustering quality fluctuations brought by randomness can not be ignored. To this end, an optimization strategy, the K-means + + initialization algorithm, is introduced to significantly improve the stability and accuracy of the clustering effect through a carefully designed initial center point selection process. K-means + + algorithm follows a probability distribution mechanism to choose the initial center, the specific process is as follows: first randomly select a data point as the first cluster center, and then for each point in the data set, calculate the distance with the nearest distance, the distance as the next center of the probability distribution [6]. In this way, the subsequent center points tend to be selected far away from the current center points, thus avoiding the problem of overconcentrated initial center points.

X C X X Formally, given the data set, the selected set of central points is, for each point in, the probability of being selected as a new center is proportional to, where is the distance to the nearest central point, i. e.

$$P(x)\ D(x)^2\ D(x)\ x\ P(x)=\frac{D(x)^2}{\sum_{x'\in X} D(x')^2} \tag{4}$$

This initialization strategy of K-means + + provides a better starting point for the iteration process by reducing the stochticity of the initial center point selection, thus effectively reduces the number of iterations, improves the overall efficiency and clustering quality of the algorithm, and provides a solid foundation for the personalized learning path recommendation based on data mining [7].

2.2 Apriori algorithm

2.2.1Apriori Introduction of the algorithm

Apriori As a classical algorithm for association rule mining, the core idea is to find frequent items set iteratively, so as to find strong rules between terms. The algorithm is based on a key premise: if a set of items is frequent, then all of its non-empty sets must also be frequent [8]. This property leads the algorithm to gradually scale up the set of terms in a hierarchical way, starting with a single term and gradually merging frequent term sets to explore larger sets of terms. In each iteration, the algorithm first calculates the support of the candidate set, selects the item set that meets the minimum support threshold as the frequency set, and then generates a new candidate set based on the current frequency set, repeating the process until a new frequent set cannot be generated. Support (Support) and confidence (Confidence) are two key indicators in the algorithm, defined as the frequency of the item set and the conditional probability of appearing under the condition of occurrence, i. e X X Y

$$Support(X)=\frac{|X\cap D|}{|D|} \tag{5}$$

$$Confidence(X\Rightarrow Y)=\frac{Support(X\cup Y)}{Support(X)} \tag{6}$$

D Which represents all transactions in the database. Despite its landmark significance in the field of association rule mining, the Apriori algorithm needs to scan the database repeatedly to calculate support when processing large-scale datasets, resulting in computational inefficiencies [9]. Moreover, the memory consumption of the algorithm during the generation of large sets of candidates also becomes an important factor limiting its application. Therefore, the improvement of Apriori algorithm aims to optimize the generation process of candidate set and reduce the number of database scans to improve the efficiency and scalability of the algorithm, especially in the application scenarios for learning path recommendation, the optimization and improvement of the algorithm is particularly important.

2.2.2 Improve the Apriori algorithm

One of the improvement strategies is to use complex data structure and algorithm optimization techniques, such as the projection-based frequent mode growth (Projected Frequent Pattern Growth, PFP-growth) technology, which improves efficiency by reducing the dimension and complexity of the data set [10].

The PFP-growth algorithm uses the concept of conditional mode base (Conditional Pattern Base, CPB) and projection database (Projected Database) to improve the FP-growth algorithm. CPB is a collection of prefix paths given a suffix pattern of frequency terms, each path from the root to a given term. By recursively processing these projection databases, PFP-growth avoids generating a large set of candidates, particularly performing well in high-dimensional data [11]. The core of the PFP-growth algorithm is to build the conditional FP-tree (conditional frequent pattern tree), and recursively mine each conditional FP-tree to find the set of frequent items. The algorithm process can be expressed as:

(1) Build the initial FP-tree.

(2) For each frequency item, its CPB is constructed, and then the projection database is obtained.

(3) Recursively build conditional FP-tree on each projection database and mine frequent item sets.

This process can be simplified by mathematical formulas, with a set of items, and a transaction in the database is a subset of the set of items. For the item set, its support count is defined as the number of transactions included. The conditional pattern is based on the set of shaped transactions, which represents the result after removing the set of items from the transaction.

$$I = \{i_1, i_2, \cdots, i_n\} \; D \; t \; I \; X \; X \; \{(t - X) : t \in D \wedge X \subseteq t\} \; t - X \; t \; X \tag{7}$$

The introduction of the improved algorithm not only significantly improves the processing power of the algorithm in the big data environment, but also maintains the high accuracy while improving the efficiency of the association rule mining. Especially for educational data analysis, improved technologies such as PFP-growth can effectively deal with complex and changeable learning data, providing strong technical support for personalized learning path recommendation [12].

3. Research Methods

3.1 Overall design

In the overall design of building a constructive English learning platform, the core goal is to achieve an efficient, dynamic and personalized learning environment by integrating advanced data mining technology —— with optimized k-means algorithm and improved Apriori algorithm ——. The architecture design of the platform follows the principle of modularity, dividing the complex system into four core modules: data processing, learner model building, content recommendation and interactive feedback, to support flexible function expansion and technology upgrading. First, the clustering analysis of massive learner data is performed through the improved k-means algorithm, using the loss function optimization, learning rate adjustment and initial center point selection strategy in the algorithm to accurately characterize learner characteristics and requirements. Subsequently, the improved Apriori algorithm is used to mine the association rules between the learning contents, and personalized learning paths are generated efficiently, based on the concept of the conditional pattern based on the projection database [13]. The data layer of the platform adopts high-performance database management system to ensure efficient data processing and safe storage; the application layer combines the latest Web technology and artificial intelligence algorithm to provide intuitive and friendly user interface and intelligent learning experience. The whole platform design fully considers the scalability and adaptability of the system, and aims to provide customized English learning services for learners with different backgrounds and needs, so as to enhance the sense of participation and satisfaction of learners while improving the learning efficiency.

3.2 Experimental environment

As shown in Table 1, in the experimental environment construction of the constructed English learning platform, the configured and efficient hardware and stable software resources are selected. The core processing unit uses a Intel Xeon Gold 6226R processor, with 16 cores and 32 threads, and the main frequency reaches 2.9GHz, ensuring the high speed and high efficiency of data processing. The system memory is 128GB DDR4 ECC, which can support a large number of concurrent data operations and complex algorithm operation. In terms of data storage, 2TB NVMe SSD hard disk is used to achieve fast read and write speed and ensure the fluency of data processing [14]. The operating system chose Linux Ubuntu 20.04 LTS, whose openness and high degree of customability facilitate development. The development language uses Python 3.8, relying on its rich data science libraries such as NumPy, SciPy, Pandas, and Scikit-learn, which provide strong support for data processing and analysis. The configuration of the overall experimental environment is designed to provide a stable and efficient technical foundation for the research and development of the constructed English learning platform, to ensure that the platform can efficiently process large-scale learning data, and support the operation of complex data analysis and machine learning algorithms.

Table 1 Table of experimental environmental parameters.

assembly	Detailed configuration
processor	Intel Xeon Gold 6226R, 16 core 32 threads, main frequency 2.9GHz
internal storage	128GB DDR4 ECC
memory	2TB NVMe SSD
operating system	Linux Ubuntu 20.04 LTS
development language	Python 3.8
Mainly rely on library	NumPy, SciPy, Pandas, Scikit-learn

3.3 Functional design

3.3.1 Personalized learner group function

(1) The user clustering mechanism based on the k-means algorithm

In the process of building a constructive English learning platform, one of the keys to realizing personalized learning experience is to accurately identify and divide learners' groups in order to provide customized learning paths. The user clustering mechanism based on the k-means algorithm is precisely designed to solve this challenge. By analyzing their behavioral data and learning preferences, the mechanism uses the k-means clustering algorithm to divide the learners into different groups. In the specific operation process, the multidimensional feature space is first defined, in which each dimension represents a specific behavior or preference index of the learner, such as learning time, course completion rate, point of interest, etc [15]. Then, according to these multi-dimensional characteristics, the k-means algorithm is used for iterative calculation. In each iteration process, the learner is dynamically adjusted according to the distance between the learner characteristics and the group center until the optimal group division is reached, that is, the algorithm converges and the group center changes small or the preset number of iteration.

(2) Application of loss function optimization in user grouping

In the core function of building the constructed English learning platform —— personalized learners group ——, the adopted strategy is realized through the refined k-means algorithm loss function optimization. This process, starting with the weighted loss function introducing data point-specific weights, is formalized as

$$J' = \sum_{i=1}^{k} \sum_{x \in S_i} w_x \left\| x - \mu_i \right\|^2 \tag{8}$$

w_x Where adjusted to attenuate the negative effect of outliers on clustering accuracy.

In parallel, the regularized term

$$R(\mu) = \lambda \sum_{i=1}^{k} \left\| \mu_i \right\|^2 \tag{9}$$

λ It is incorporated into the loss function, aiming to balance the model complexity, avoid the overfitting phenomenon, and enhance the generalization ability of the algorithm. During the iteration process of the algorithm, the clustering center is constantly adjusted according to the weighted loss and regularization condition until the convergence standard is reached. At this time, the clustering results reflect the detailed division of learners in behavior and preference. The application of the optimization strategy ensures the high efficiency and accuracy of the algorithm in the face of multi-dimensional learner data, lays a solid foundation for the construction of accurate personalized learning path, and shows the ability to deeply analyze and meet the personalized needs of education at the technical level [16].

(3) The effect of the learning rate and the initial center point selection on the grouping effect

In the personalized learner clustering function of the constructed English learning platform, the learning rate of k-means clustering algorithm and the K-means + algorithm for the initial center point selection realize highly accurate and efficient user clustering. Learning rate was adjusted using an adaptive approach, with the initial learning rate set at 0.1 and decayed exponentially with the number of iterations.

The specific formula is that,

$$\alpha_t = \frac{\alpha_0}{1 + \delta t} \tag{10}$$

α_0 δ t Where is the initial learning rate, which is the decay rate, and was set to 0.05 in this experiment, which is the current number of iterations. This dynamic learning rate strategy enables the algorithm to quickly approach the optimal solution in the early stage and reduce the update step size near convergence, thus avoiding excessive oscillations and improving clustering stability and accuracy [17].

For the initial center point selection, the K-means + + algorithm was applied to optimize the starting clustering center and further improve the quality of the clustering effect. K-means + + gradually selects the initial center point through a probability distribution strategy to ensure that the newly selected centers are as far away from the selected ones. The first step of this strategy is to randomly select a data point as the first center point. Next, the probability that each unselected data point is selected as the next center point is proportional to the square of its distance to the nearest selected center point.

This probability is calculated as follows,

$$P(x) = \frac{D(x)^2}{\sum_{x' \in X} D(x')^2} \tag{11}$$

$D(x)$ x X Where is the distance from the point to the nearest center, which is the set of all points in the dataset. This approach reduces the randomness of the initial center point selection, making the clustering results more stable and reliable.

3.3.2 Dynamic Learning path recommendation system

(1) Improve the application of the Apriori algorithm in the learning path recommendation

In the dynamic learning path recommendation system of constructing a constructed English learning platform, the core of the improved Apriori algorithm is applied to mine the deep correlation rules between learning contents, so as to provide a personalized learning path for learners to customize. The algorithm achieves efficient recommendation of learning paths by finely tuning the minimum support and minimum confidence parameters and using advanced data structure to optimize the rule generation process.

In terms of specific technical implementation, the improved Apriori algorithm sets the minimum support threshold of 0.05 and the minimum confidence threshold of 0.3, which aims to balance the universality and pertinence of the rules and ensure that the mined association rules are both universal and practical. Moreover, by introducing the transaction compression technique and the hash tree structure of frequent term sets, the algorithm is able to greatly reduce the computational amount in the process of trathrough the database for frequent terms sets, improving the ability to process large-scale learning data [18].

In the process of mining the association rules between learning contents, the improved Apriori algorithm first builds a transaction database based on the learner's historical behavior data, and each transaction record represents a learner's learning activity sequence. Applying the set minimum support and confidence parameters, the algorithm iteratively identifies frequent combinations of learning activities and their dependencies on each other. Through this process, the algorithm can reveal the key learning path information, such as the order between specific learning materials and the complementary relationship between different learning activities.

(2) Association analysis of learning content

In the dynamic learning path recommendation system of the constructed English learning platform, the learning content correlation analysis function accurately excavates the deep correlation between learning materials by using the improved Apriori algorithm, providing scientific basis for learning path design. This process is centered on the improved Apriori algorithm, and ensures that association rules with both frequency and high confidence are extracted from a large amount of learning activity data by setting refined technical parameters such as the minimum support threshold of 0.04 and the minimum confidence threshold of 0.2.

First, the initial set of transactions is constructed based on the learner interaction data, and each transaction represents a learner's pattern of learning behavior in a specific time period. Through the improved Apriori algorithm, the frequent item sets are filtered and generated layer by layer. In this process, efficient data structure and algorithm optimization techniques are introduced, such as transaction hashing and item set hashing techniques, which greatly reduces the computational complexity and improves the operation speed. Subsequently, for each set of frequent items, its confidence with the other item sets is calculated to generate association rules. These rules reveal dependencies between different learning materials, such as the order of specific course content, or the auxiliary role of one learning resource in understanding another resource. Further, the learning path design is optimized by evaluating the advanced indicators such as the improvement (lift) (conviction) and certainty of the association rules. The improvement index reflects the relevance of the rules, while the certainty provides the reliability evaluation of the rules. Combined with these indicators, the platform can screen out the most influential and guiding learning content association rules [19].

(3) Real-time adjustment mechanism of the recommendation system

In the dynamic learning path recommendation system of the constructed English learning platform, an advanced real-time adjustment mechanism is introduced, which can dynamically adjust and optimize the recommended learning path according to the immediate feedback and learning progress of learners. The real-time adjustment mechanism is based on a set of refined algorithm formulas and parameter settings to ensure the maximum learning efficiency and effect. Central to this mechanism is the feedback adjustment function

$$F(R,P)=\alpha R+(1-\alpha)P \tag{12}$$

R P α It represents the immediate feedback score of the learners, which represents the progress score of the learners, and is the balance coefficient, which is used to adjust the weight of the influence of the immediate feedback and the learning progress on the learning path adjustment. In practice, it is usually set to 0.5 to ensure an equal impact of immediate feedback and learning progress on learning path adjustment. α

R P The immediate feedback score is obtained by the learners' satisfaction survey or direct evaluation of the learning content, while the learning progress score is calculated by comparing the course units completed by the learners with the number of predetermined course units on the recommended path. The specific formula is

$$P = \frac{N_C}{N_t} \tag{13}$$

N_C N_t $F(R,P)$ $F(R,P)$ This is the number of completed course units and the total number of course units on the recommended path. According to the calculation results of the feedback adjustment function, the system will update the learners' learning path in real time. If the value is lower than the preset threshold (e. g., 0.7), the system will trigger the recalculation and adjustment process of the learning path, introducing learning materials that more meet the needs and progress of learners. At the same time, the subsequent content on the learning path may be reordered or replaced to optimize learning efficiency and improve learning effectiveness.

3.3.3 Interactive learning feedback system

(1) Learner behavior data collection

In the constructed English learning platform, the interactive learning feedback system captures the behavioral data of learners through the refined data collection mechanism, providing rich basic information for the subsequent data analysis and learning path optimization. The system adopts a multi-dimensional data collection method, covering key indicators such as learning time, frequency, and ensures the comprehensiveness and accuracy of the data by setting precise technical formulas and parameters [20].

Learning time data collection relies on the timestamp recording function

$$T(s,e) = e - s \tag{14}$$

s e Where and represent the beginning and end time points of the learning activity, respectively. This function can accurately calculate the total time spent by learners in each learning activity, providing quantitative indicators for the analysis of learning investment.

The learning frequency is measured by the frequency calculation formula

$$F = \frac{N_d}{\triangle t} \tag{15}$$

N_d $\triangle t$ To achieve, where, represents the total number of learning activities during the time period. The formula helps the platform to monitor learners' learning activity, reflecting their learning habits and patterns.

The capture of learning preferences, then combined with content interaction analysis,

Through the scoring function

$$P(c_i) = \frac{\sum_{j=1}^{n} r_{ij}}{n} \tag{16}$$

r_{ij} c_i n To achieve, which is the learner's score of the first learning content, is the number of scores. By calculating the average score, the system can identify the preferred learning content and form of learning, and then provide the basis for personalized recommendation.

Advanced data analysis techniques such as machine learning algorithms are used to conduct an in-depth analysis of the collected data to identify learners' behavior patterns and potential needs [21].
Applied the cluster analysis method

$$C = \arg\min_{c \in C} \sum_{i=1}^{k} \sum_{x \in S_i} \left\| x - \mu_i \right\|^2 \tag{17}$$

Learners were grouped according to the behavioral data to identify different groups of learners and their characteristics.

(2) Personalized feedback based on the data analysis

In the interactive learning feedback system of the constructed English learning platform, the personalized feedback function based on data analysis adopts the k-means algorithm to conduct in-depth analysis of the behavioral data of learners to provide targeted learning feedback. This process involves precise algorithm parameter configuration and data processing techniques that ensure accurate identification of learners' learning patterns and potential needs.

In the specific implementation, the multi-dimensional feature space is first defined to represent learners' behavioral data, including but not limited to indicators such as learning time length, learning frequency, course completion degree and interactive feedback. Based on these dimensions, applying the k-means algorithm for cluster analysis,

Formulation to solve the minimization problem,

$$J = \sum_{i=1}^{k} \sum_{x \in S_i} \left\| x - \mu_i \right\|^2 \tag{18}$$

S_i μ_i k k Where represents the set of learners in the first cluster, which is the center of the cluster. By setting a reasonable number of clusters (such as determining the optimal value by the elbow method), the algorithm is able to divide learners into different groups according to their behavioral characteristics [22].

After the cluster analysis, the platform identified the learning blind spots and dominant regions of learners by comparing the deviation of the learning behavior characteristics with the cluster center. For example, if a learner is well below the center of the cluster in the length of learning time, the system can identify this as a potential learning blind spot.

Subsequently, the platform generates personalized learning feedback reports based on these analysis results. The report not only points out the learning blind spots of learners, but also provides targeted suggestions and optimizing learning strategies, such as adjusting learning plans, recommending additional learning resources or practical activities, aiming to help learners comprehensively improve their learning efficiency [23].

In addition, the platform also introduces a dynamic feedback mechanism to update the personalized feedback content in real time according to the learners' response to the early feedback and the changes in learning behavior. This mechanism ensures the real-time nature and adaptability of the feedback, enabling learners to be effectively guided and supported in the ongoing learning process.

4. Application

4.1 Application environment and parameter setting

As shown in Table 2, the hardware environment uses an Intel Xeon E5-2620 v4 processor with 64GB RAM, ensuring sufficient computing resources and memory to support large-scale data processing. The software environment is based on the Linux Ubuntu 18.04 LTS operating system, combined with Python 3.7 and its data processing and machine learning libraries, such as NumPy, Pandas, and Scikit-learn, providing an efficient and stable platform for algorithm implementation and data analysis. The k-means algorithm and the improved Apriori algorithm parameters are finely adjusted to balance the efficiency of the algorithm with the accuracy of the mining results, in which the cluster number of the k-means algorithm is determined by the elbow rule, while the minimum support and confidence thresholds of the improved Apriori algorithm are optimized to accommodate the learning data characteristics. In addition, the data set covers the learning behavior and content data, comprehensively reflecting the learner interaction and learning material characteristics. Through this series of configurations and Settings, the experiment aims to accurately explore and evaluate the effectiveness of the learning path recommendation system, and further enhance the personalized learning experience and educational value of the learning platform.

4.2 Application index setting

In this study, a series of refined test indicators were used to deeply explore the performance and effects of the core functions of the constructed English learning platform —— "Personalized learner group function", "Dynamic Learning path recommendation system" and "Interactive Learning feedback system" ——. As shown in Table 3, these indicators cover key dimensions such as clustering homogeneity, clustering stability, recommendation accuracy, and user satisfaction, and comprehensively evaluate the technical performance of each function and its impact on user experience. Through a detailed analysis of cluster separation, within-group cohesion, recommendation relevance, diversity, and other indicators, this study can accurately assess the efficiency and effect of personalized learners' clustering, as well as the accuracy and user acceptance of the dynamic learning path recommendation system [24]. At the same time, the setting of feedback timeliness, personalization degree and learner participation provides an important basis for evaluating the real-time, personalized service ability and learning effect improvement of the interactive learning feedback system. Furthermore, the introduction of common metrics such as retention and overall satisfaction further ensures a comprehensive assessment of the overall performance of the platform's performance and user satisfaction. In conclusion, the careful design and application of these comprehensive trial metrics provide a solid data support and scientific evaluation basis for optimizing the functional realization of learning platforms, improving the user learning experience and promoting the practice of promoting personalized learning [25].

Table 2 Table of experimental parameters settings.

class	CI	Detailed configuration
hardware environment	processor	Intel Xeon E5-2620 v4(2.1GHz, 8 Core)
	internal storage	64GB RAM
software environment	operating system	Linux Ubuntu 18.04 LTS
	programming language	Python 3.7
	Data processing library	NumPy, Pandas
	Machine learning library	Scikit-learn
	Visualization library	Matplotlib
parameter setting	k-means algorithm	
	Number of clusters, k k	5 to 15 (elbow rule)
	Initialization method	K-means++
	Maximum iteration times	300
	distance measure	Euclidean distance
	Improved Apriori algorithm	
	Minimum support threshold	0.03
	Minimum confidence threshold	0.2
	Minimum lift degree threshold	1.5
data set	Learning behavior data	Learning time, learning frequency and course interaction
	Learn content data	Course structure, learning materials, test questions and other relevant information

4.3 Analysis of the application results

4.3.1 Analysis of the experimental results of personalized learners' group function

In the personalized learner group function experiment of this study, the results of the test group and the control group showed that the improved k-means algorithm significantly improved the homogeneity and separation of the clusters. As shown in Table 4, the cluster homogeneity of the test group reached 0.87, which was significantly improved compared with the 0.75 of the control group, indicating that the behavior pattern and learning preferences of the learners within the same cluster were more consistent. Meanwhile, the clustering stability index reached 0.92, which was higher than the 0.80 of the control group, indicating that the clustering results had higher consistency and reliability on different data sets. In terms of the running time of the algorithm, the test group only took 2.3 seconds to complete the group processing of the learners, which significantly decreased compared with the 3.5 seconds of the control group, showing the significant effect of the optimization algorithm in improving the data processing efficiency. In addition, the improvement of cluster separation and intra-group cohesion reached 0.82 and 0.89, respectively, which further confirmed that the algorithm can maintain close contact within the group, providing a solid foundation for accurate personalized recommendation."Group coverage" and "group feature recognition accuracy" index, respectively showed 95% coverage and 90% recognition accuracy, which shows that the improved algorithm can not only cover a wider group of learners, also can accurately identify the specific characteristics of learners, further enhance the accuracy and effectiveness of personalized learning recommended.

Table 3 Schematic diagram of the test indicators.

Core function	test index	description
Personalized learner group function	Cluster Homogeneity (Homogeneity)	Measure whether learners' behavior and preference are highly consistent with other members within the same cluster.
	Group stability (Stability)	To evaluate the consistency of cluster results on different data subsets and measure the stability of the algorithm under different data samples.
	Algorithm running time (Algorithm Runtime)	The total time to complete the learner group was recorded to reflect the processing efficiency of the algorithm.
	Cluster separation degree (Separation)	Measure discrimination between clusters, high separation indicates a clear distinction between clusters.
	Intragroup cohesion (Cohesion)	To assess the closeness between members within the same cluster, the high cohesion indicates similar characteristics among members.
Dynamic learning path recommendation system	Recommended accuracy rate (Accuracy)	Measure the agreement of the systematically recommended learning path with the actual needs and preferences of the learners.
	Recommended Response Time (Response Time)	Average time from the user request the recommendation to the system generating the recommendation result.
	User satisfaction (Satisfaction)	Users' satisfaction with the recommended learning path was collected through a questionnaire survey or an online feedback mechanism.
	Recommended Correlation (Relevance)	The relevance of the recommended content was measured by comparing with the learner's actual learning goals and preferences.
	Recommended Diversity (Diversity)	To evaluate the diversity of courses and materials in the recommended learning path to ensure that learners are exposed to a wide range of knowledge points.
	Recommended stability (Stability)	Analyze the consistency of the system recommendation results when the same learner requests the recommendation at different times.
	Novelty (Novelty)	To assess how much unknown or unexpected learning content recommendation systems can offer to learners.
Interactive learning and feedback system	Feedback timeliness (Timeliness)	Feedback generation with speed of delivery to learner learning motivation and feedback adoption rate.
	Degree of feedback (Personalization)	Evaluation of whether the feedback content is tailored to the specific needs and learning status of the learners, reflecting the personalized service capability of the system.
	Feedback accuracy (Precision of Feedback)	Measure how well the learning feedback provided by the system fits with the learners' actual needs and problems.
	Learner Engagement (Learner Engagement)	It is measured by learners' interactive behavior towards feedback (such as click rate, completing feedback suggestion behavior, etc.).
	Learning Effect Improvement Index (Learning Improvement Index)	Compared with before and after receiving personalized feedback, learners improved their performance on specific learning objectives and quantified the improvement of learning effect.
General indicators	User Retention Rate (User Retention Rate)	The proportion of learners consistently using the system for a certain period of time reflects the overall attractiveness and user engagement of the system.

	Overall satisfaction rate (Overall Satisfaction)	The satisfaction scores of the whole platform obtained through the comprehensive survey directly reflects the overall level of service quality and user experience of the system.

4.3.2 Analysis of the experimental results of the dynamic learning path recommendation system

Results of the experimental evaluation of the dynamic learning pathway recommendation system reveal significant advantages of the system on key performance metrics. As shown in Table 5, in terms of recommendation accuracy, the results in the test group reached 84%, a significant increase of 16 percentage points compared with 68% in the control group. This significant improvement is due by the improved recommendation algorithm, which analyzes the behavioral data and preferences of learners more deeply, so as to more accurately predict and meet the learning needs of learners. The optimization results of the recommended response time were also impressive, with the average response time in the test group being 1.5 seconds, which significantly shortened and almost decreased by half from the 3.0 seconds in the control group. This improvement directly impacts the fluency of the user experience, ensuring that learners can quickly access the required learning resources to seamlessly continue their learning process. The increase in user satisfaction to 90%, 12 percentage points higher than the control group, reflects the high recognition of the overall performance of the recommendation system and the quality of the recommended content. This result reflects the success of the recommendation system in meeting the personalized needs of users and providing high-quality learning content. The improvement of recommendation diversity and novelty reached 82% and 85% respectively, indicating that the recommendation system not only ensures the accuracy of recommendation, but also successfully provides users with a rich variety of learning content, which effectively stimulates learners' interest and motivation in learning.

Table 4 Results of experiments on personalized learners' cluster function.

name of index	Test group results	Results of the control group	remarks
Cluster homogeneity	0.87	0.75	markedly improve
Group stability	0.92	0.8	Higher stability
Algorithm run time	2.3 Seconds	3.5 Seconds	Significantly lower
Clustering separation	0.82	0.7	The separation degree is better
Cohesion within the group	0.89	0.78	The cohesion is stronger
Group coverage	95%	85%	Increased coverage
Group feature identification accuracy	90%	80%	Accuracy improvement

Table 5 Results of the dynamic learning pathway recommendation system experiments.

name of index	Test group results	Results of the control group	remarks
Recommended accuracy	84%	68%	Improve by improving the algorithm
Recommended response time	1.5 Seconds	3.0 Seconds	The response is accelerated after optimization
User satisfaction	90%	78%	The user experience has been significantly improved
Recommend diversity	82%	70%	The content is much richer and more diverse
Recommended stability	87%	75%	System recommendations are more stable
novelty	85%	65%	The novelty degree of the recommended content is improved

4.3.3 Analysis of experimental results of interactive learning feedback system

In the evaluation of the interactive learning feedback system, by comprehensively examining the key indicators such as feedback timeliness, personalization degree, accuracy and learners 'participation, the system shows the deep understanding and rapid response ability of learners' behavior and needs. As shown in Table 6, the test group only needed 3.5 seconds on average of feedback timeliness to generate and provide feedback. Compared with the 5.5 seconds in the control group, it significantly shortened the waiting time of learners and accelerated the learning cycle, thus improving the learning efficiency. The degree of personalized feedback has reached 92%, which is significantly improved compared with 78% of the control group, indicating that the interactive learning feedback system can provide more customized learning suggestions and resources according to the specific situation and needs of learners, and strengthen the personalized learning experience. In addition, the improvement in the accuracy of feedback to 89% reflects the system's accurate grasp and effective guidance of learners' problems and puzzles, ensuring the practicability and pertinacity of feedback. The participation of learners increased to 87%, 18 percentage points higher than the control group, indicating the high recognition and positive response of learners to the system feedback. This increase in engagement not only reflects the attractiveness and motivating effect of the system feedback, but also indirectly promotes the learning motivation and learning effectiveness of the learners. The learning effect improvement index reached 1.8, compared with 1.2 of the control group, significantly demonstrating the effectiveness of interactive learning feedback in promoting learners' understanding and mastering the learning content [26]. Through the personalized, accurate and timely feedback received, learners can quickly identify and make up for the learning blind spots, and optimize the learning strategies, so as to significantly improve the learning results.

Table 6 Experimental results of the interactive learning feedback system.

name of index	Test group results	Results of the control group	remarks
Feedback timeliness (seconds)	3.5	5.5	The response is accelerated
Feedback degree of personalization (%)	92%	78%	significantly enhance
Feedback accuracy is (%)	89%	72%	Precision improvement
Learner Engagement (%)	87%	69%	Increased participation
Learning effect improvement index	1.8	1.2	The effect has been significantly improved

4.3.4 Results analysis of general indicators

In the general indicator evaluation of the constructed English learning platform, the significant increase in user retention and overall satisfaction highlights the success of the platform in maintaining long-term user engagement and providing a high-quality learning experience. As shown in table 7, the user retention rate rose to 85%, compared with the control group increased by 10%, the results show that the platform through its personalized learners group function, dynamic learning path recommendation system and the integrated application of interactive learning feedback system, effectively promote the user's continuous participation, enhance the user loyalty to the platform. The overall satisfaction rate increased to 90%, up by 10 percentage points compared with the control group, further proving the advantages of the platform in meeting users' learning needs and improving learning efficiency. This significant increase in satisfaction reflects users' high recognition of the quality of the learning content provided by the platform, the degree of personalization of the learning path, and the timely and effective learning feedback.

Table 7 Results of general indicators.

name of index	Test group results	Results of the control group	remarks
User retention rate is (%)	85%	75%	markedly improve
Overall satisfaction with (%)	90%	80%	significantly enhance

5. Conclusion

This study deeply explores the influence of the core functions within the construction of English learning platform —— personalized learners' subgroup, dynamic learning path recommendation system, and interactive learning feedback system —— on promoting personalized learning. The results show that, through algorithm optimization and data analysis, the platform can significantly improve the learning efficiency, meet the personalized needs of learners, and enhance the personalized and interactive learning experience. The personalized learners 'grouping function provides a solid foundation for the recommendation of dynamic learning path by improving the accuracy and efficiency of learners' grouping. After optimization, this function not only improves the homogeneity and cohesion of the group, but also guarantees the real-time response ability of the system by reducing the running time of the algorithm, and effectively promotes the personalization and accuracy of learning path recommendation. The optimization of the dynamic learning path recommendation system makes the recommended learning path more in line with the actual needs and preferences of learners, and significantly improves the recommendation accuracy and user satisfaction. Through the introduction of novelty and diversity indicators, the system can provide learners with a more extensive and more exploratory learning content, effectively stimulate learning motivation and interest, and improve learning efficiency and effectiveness. The implementation of interactive learning feedback system greatly improves the timeliness and personalization degree of learning feedback, and ensures the timeliness and relevance of feedback. System accurate analysis of learners' behavior can provide targeted learning strategies and suggestions, help learners adjust learning methods in time, and effectively improve learning effectiveness.

In summary, the analysis of the core technical functions of the constructed English learning platform demonstrates the validity of its effectiveness in supporting personalized learning. The integrated application of these functions not only meets the personalized needs of learners, but also significantly improves the learning efficiency and results, and provides a solid technical support for personalized learning. Future studies need to further explore the algorithm optimization strategies and the application effects of these functions in different learning environments, in order to deepen the understanding of building efficient and personalized learning environments.

Acknowledgments

This work was supported by the Key research projects in higher education in Jilin Province. Project Name: Research on the Professional Development Path of English Teachers in Higher Education Institutions in Jilin Province Based on Corpus Data Driven Learning, Project Number JGJX2023C123.

Reference

[1] Ma Kun, Wang Zhe, Fan Wenbo. Precision funding approach for the student aid system based on the improved Apriori algorithm. Journal of Jilin University (Engineering Edition), 2023, 53 (11): 3246-3252.
[2] Xu Qiang, Wang Shzuo. Big data AR mining simulation based on an improved Apriori algorithm. Computer Simulation, 2023, 40 (07): 509-513.
[3] Qiu Huizhi. Study of the Apriori improvement algorithm based on the vector matrix. Journal of University of Shanghai for Science and Technology, 2022, 44 (01): 56-61 + 68.
[4] Gu Haiyan, Zheng Qiwen. Research on the FT-Tree improvement algorithm for log template extraction. Nanjing Normal University Daily (Natural Science edition), 2021, 44 (02): 121-126.
[5] Sun Ningna, Li He. Design strategy of middle school English online learning interface based on persuasion design. Packaging Engineering, 2020, 41 (18): 242-248 + 279.
[6] Liu Yingchuan. Design and implementation of university educational administration system based on Apriori algorithm. Modern Electronic Technology, 2020, 43 (14): 52-54 + 58.
[7] Zhang Lei, Dong Wanfu, Kan Welcome, et al. Customer demand data analysis method based on the improved Apriori algorithm. Mechanical Design and Manufacturing, 2020, (05): 185-188.
[8] Li Jie, Zhu Hongliang, Chen Yuling, et al. A parallel Apriori improvement algorithm based on hash storage and transaction weighting. Computer Engineering, 2020, 46 (11): 109-116.
[9] Wang Qingsong, Jiang Fushan, Li Fei. Multi-label learning algorithm based on association rules in a big data environment. Computer Science, 2020, 47 (05): 90-95.
[10] Liao Jiyong, Wu Sheng, and Liu Ailian. Improvement study of the Apriori algorithm based on Boolean matrix parsimony. Computer Engineering and Science, 2019, 41 (12): 2231-2238.
[11] Wenwu, Guo Youqing. The Apriori algorithm combined with genetic algorithm. Computer Engineering and Design, 2019, 40 (07): 922-926.

[12] Zeng Yi, Zhou Xiangzhen. Improved Apriori mining algorithm based on Hadoop framework in the big data environment (English). Machine Tool and Hydraulic Pressure, 2019, 47 (06): 98-103.
[13] Gautenwen. Design of the library and information data mining and processing platform. Modern Electronic Technology, 2019, 42 (08): 25-28.
[14] Ma Jing, Qiu Guohua. Design and implementation of English information sharing platform under Ubiitous learning environment. Modern Electronic Technology, 2018, 41 (10): 93-96.
[15] Liu Lijuan. Research and application of the improved Apriori algorithm. Computer Engineering and Design, 2017, 38 (12): 3324-3328.
[16] Tan Tiechun, Han Jianping. Design of an open English learning system based on the B / S structure. Modern Electronic Technology, 2017, 40 (24): 62-64.
[17] Qu Shiqi, Liu Shaojiang, Ni Wei Chuan, et al. Accelerated Apriori algorithm for GPU clustering based on the Hadoop platform. Computer Engineering, 2018, 44 (11): 14-18 + 26.
[18] Shao Xiuli, Yin Jizhou, Zhang Huadong, et al. Integrated implementation of the K-Means algorithm on a public service platform for data analysis. Journal of Nankai University (Natural Science Edition), 2017, 50 (04): 41-45.
[19] Jin Yan, Yang Mingtuo. Design of a virtual learning space for College English based on the Virtools development platform. Audio-visual Teaching of Foreign Language, 2017, (04): 17-22.
[20] Zhao Yue, Ren Yonggong, Liu Yang. Research on the improved Apriori algorithm based on MapReduce and its application. Computer Science, 2017, 44 (06): 250-254.
[21] Jia Kebin, Li Hanjing, Yuan Ye. Application of data mining based on Apriori algorithm in mobile health systems. Journal of Beijing University of Technology, 2017, 43 (03): 394-401 + 322.
[22] Zhang Xiang. Design and practice of graphic interactive visual learning platform for thematic English reading teaching. China Audio-visual Education, 2016, (08): 125-128.
[23] Wang Ying. Strategy for improving the application ability of information technology. Exploration in Education, 2015, (08): 128-130.
[24] Yuan Xiang. Design curriculum informatization framework based on student creation content. Decoration, 2014, (08): 78-79.
[25] Zhao Weizhong, Ma Huifang, Fu Yanxiang, et al. Design of parallel k-means clustering algorithm based on cloud computing platform Hadoop. Computer Science, 2011, 38 (10): 166-168 + 176.
[26] Zhang Guanglu. Design of supporting learning environment under digital platform —— Analysis and study of class example: Necklace. China Audio-visual Education, 2006, (02): 63-65 + 71.