Data Mining of English Language Instructional System Based on Improved Collaborative Filtering Algorithm

Furong Shi[1] and Mengmeng Zhang[2]

[1]Department of Basic, Shaanxi Fashion Engineering University, Xi'an, Shaanxi 712046, China, shifr@snfeu.edu.cn
[2]Department of Basic, Shaanxi Fashion Engineering University, Xi'an, Shaanxi 712046, China, mengbin7g@163.com

Corresponding author: Furong Shi, shifr@snfeu.edu.cn

Abstract. The use of computer-aided language learning (CALL) in practical teaching has built an interactive platform for the integration of computer and language teaching. In order to solve the problem of data sparseness in the information recommendation model of English resource database with large data volume, this article proposes a data mining (DM) in database (KDD) model of CALL system based on improved CF algorithm. Through CF algorithm and the expression of resource feedback matrix, the resource selection is realized. The results show that the algorithm has high recommendation accuracy and time complexity. The model algorithm has high recommendation accuracy and efficiency, which effectively solves the problem of large data sparseness, can further enhance the effect of computer-aided instruction (CAI) system, improve the classroom experience of educators and learners, and promote the intelligent growth of CAI system.

Keywords: Collaborative Filtering; Computer-Aided Instruction; Data Mining; Knowledge Discovery
DOI: https://doi.org/10.14733/cadaps.2024.S10.136-150

1 INTRODUCTION

The social demand for English talents is growing rapidly, and people are increasingly demanding to learn English. However, the traditional English language teaching (ELT) has been difficult to fully meet the demands of the growth, and the problems exposed have attracted widespread attention from the English community and society. Advanced artificial intelligence and machine learning technologies aim to automate the testing process, improve testing efficiency, reduce human errors, and provide personalized feedback. This system can improve exam efficiency and accuracy through automation and intelligence, reduce the impact of human factors on exams, and provide more reliable and effective support for educational evaluation and teaching. The system can automatically generate test papers that meet different exam standards based on exam requirements and knowledge points. This can greatly improve the quality and difficulty of the test paper, while also reducing errors and omissions in manual preparation. Alghamdi et al. [1] achieve accurate assessment of candidates' knowledge and skills by automatically evaluating and analyzing their answers. The arrival of the computer network era has changed the traditional instructional methods, especially ELT. ELT no longer just stays in the exam-oriented education stage of memorizing words, doing problems and taking exams, but give priority to mastering the English language itself, and the practicality of the language is paid attention to. The English translation intelligent system using an improved hybrid recommendation algorithm to automate the knowledge base is a complex system. The computer-aided intelligent examination system improves the accuracy and efficiency of English translation by combining advanced technologies and methods, while providing a user-friendly interface that allows users to easily obtain the information they need. This provides users with better translation services and experiences. Bi [2] stores and queries translation information through a vast knowledge base. This knowledge base can include various types of translation information, such as fixed phrases, commonly used sentences, professional terms, etc. The system can continuously collect user feedback and adjust recommendation and translation strategies based on feedback to continuously improve the quality of translation and user satisfaction. CALL in the network age is an important method to guide students to learn English actively, which will help students greatly in learning English. Computer network can improve students' learning enthusiasm in class, and students can also actively study independently after class. In teaching, CAI system can conveniently display perceptual materials and create the best situation, thus obtaining the best instructional effect. Chen and Huang [3] utilize IoT technology to collect real-time data on students' learning behavior and grades, such as learning time, progress, practice scores, etc., as well as sensor data related to English, such as voice intonation, mouth movements, etc. These data can be transmitted to the cloud through IoT devices for processing and analysis. The system adopts interactive teaching methods, such as online situational dialogues, interactive games, virtual role-playing, etc., to enhance interaction and communication between students and the system. This teaching method can increase students' participation and learning interest, while also improving their practical application ability in English. The computer-aided intelligent examination system adopts scientific programming methods to conduct in-depth analysis and mining of the collected data, in order to understand students' learning situation and problems. By analyzing and mining a large amount of data, the system can obtain various learning behavior and performance data related to students, such as accuracy rate, time, and number of repeated answers. These data provide valuable information for the system, enabling it to more accurately evaluate students' learning situation and problems.

Chen and Deng [4] analyzed personalized recommendation of online video English learning resources. Its purpose is to recommend courses and resources that best meet learners' needs and interests through the use of data analysis and machine learning techniques. Common matching algorithms include content-based recommendation algorithms, collaborative filtering recommendation algorithms, hybrid recommendation algorithms, etc. In short, exploring personalized recommendation methods for online video English learning resources can improve learners' learning efficiency and quality. This requires a systematic analysis of the characteristics of learners and resources, and the use of data analysis and machine learning techniques for matching and recommendation. At the same time, it is necessary to continuously optimize and improve recommendation algorithms to meet the constantly changing needs and interests of learners. The recommendation algorithm based on user behavior analysis is a common personalized recommendation method. This method analyzes the learning behavior of learners, such as the duration of watching videos, the number of repeated views, likes, etc., to understand their interests and needs, and recommend relevant video resources for them. It can accurately recommend based on the personalized needs of learners, improving the success rate of recommendations. This method can use different machine learning algorithms, such as collaborative filtering, content filtering, deep learning, etc., to improve the accuracy and personalization of predictions. The advantage of machine learning based recommendation algorithms is that they can handle complex user interests and video features, improving recommendation accuracy. The application of CALL in practical teaching combines the means of CALL with this new curriculum concept, realizes the internalization and externalization of educational ideas and concepts, and builds an interactive platform for the integration of computer and language teaching. The vast majority of existing intelligent teaching systems are running in a stand-alone environment, and it is impossible to take advantage of the convenience and rapidity of the network in updating knowledge. Such a system cannot update and maintain the instructional content, and provide more targeted teaching for students. The course CAD system realizes the recommendation function to meet the demands and interests of users by mining. The birth of course CAD system can solve the problem of online course information explosion. Developers should develop highly reliable course recommendation algorithms, quickly and accurately obtain the course information needed by users, and give accurate recommendation results, which has become the research focus. The main content of this article is to propose a individualized course recommendation method for students' hobbies, and construct the DM and KDD model of the CALL system.

The increasing abundance of educational resources, finding suitable English education resources from massive amounts of information is an important challenge for both students and teachers. Fan [5] explored a data mining-based method for recommending English education interest texts, aiming to provide readers with an effective means to discover and recommend English education texts related to their interests. Data mining is the process of extracting information and knowledge from massive data, which can be used to discover patterns and relationships in the data. In English education interest text recommendation, data mining can help identify readers' interest areas, search for and recommend relevant texts. The English education interest text recommendation method based on data mining can effectively discover and recommend English education texts related to readers' interests, improving the effectiveness and efficiency of English education. In the future, we will continue to research and explore more advanced recommendation technologies and methods to provide higher quality services for English education. Huge and complicated data will make it difficult for people to choose, so it is needed to establish an information recommendation model of relevant English databases according to the user's interest in the project, select the most suitable resources and provide effective information recommendation services for them. Automatic course recommendation is becoming the key function of course selection system and course learning platform, and recommendation algorithm plays a key role in the course recommendation function. In this article, DM and KDD model of CALL system based on improved CF algorithm are constructed:

⊖ The model deeply digs the historical monitoring data, transforms the original data into a resource data set through integration and perfection, and fits the overall similarity threshold of the project on this basis.

⊜ The algorithm further controls the decay rate of the time function by introducing the user interest curve and the time decay factor. At the same time, the average scoring method is used to capture users' recent interests, so as to effectively grasp the changes of users' interest preferences.

The content structure of this article is arranged as follows:

The first section introduces the significance of the construction of CALL system and the development needs of individualized recommendation model; In the second section, the DM and KDD model based on the improved CF algorithm are constructed. In the third section, the recommendation performance of the model is verified by simulation experiments. The fourth section summarizes the main work and achievements of this article and puts forward the possible improvement direction.

2 RELATED WORK

Gong et al. [6] utilized the advantages of computer-aided technology and corpus management to provide learners with a highly personalized, interactive, and practical learning environment. The core of the system is a corpus based on computer-aided technology, which contains a large amount of real English corpus, including articles, sentences, dialogues, movie scripts, etc. This corpus not only covers various themes and fields, but also includes multiple languages and dialects. This system achieves automated and intelligent teaching management. For example, by analyzing students' learning records and test scores, personalized learning plans and assignments are automatically generated, and suitable learning resources and strategies are intelligently recommended based on students' learning progress and needs. Based on corpus management, provide rich and authentic English learning materials. The system collects a large amount of English corpus resources, including text, audio, video, etc. It can provide diverse scenarios according to different learning needs and teaching objectives, helping students better master English language skills. Create a comprehensive English learning environment by combining online learning and offline practice. This system not only provides online courses and learning resources, but also interfaces with offline English learning platforms or physical classrooms to achieve online and offline learning interaction and communication. Huang and Zhu [7] utilized corpus and natural language processing techniques to construct a knowledge graph in the field of English textbooks. This system can recommend the most suitable English textbooks for learners based on their personalized needs and characteristics. This system not only improves learners' learning efficiency and quality, but also provides better teaching aids for teachers, promoting the development of English teaching. It can provide accurate textbook recommendation services based on the personalized needs of learners, which helps to improve English learning effectiveness and enhance learner satisfaction. In the future, we will continue to research and optimize the system to provide stronger support for personalized learning. By using natural language processing and machine learning techniques, key information is extracted from a large number of English textbooks to construct a personal knowledge graph. By combining machine learning and collaborative filtering algorithms, personalized English textbooks are recommended for learners based on their personal information and historical learning behaviors. Based on learners' feedback and learning behavior, update personal knowledge graph in real-time to improve recommendation accuracy. Developing a computer-aided English teaching system using Visual Basic (VB) can greatly improve teaching efficiency and enhance students' learning experience. The following is a simple step-by-step guide to help you design and implement such a system using VB. Jing and Jiang [8] use VB's graphical user interface design tools to create user interfaces. This interface should be easy to use and provide clear information. Write backend logic code based on user interface requirements. For example, when a user clicks the submit job button, the code should be able to retrieve job information and save it to the database. Conduct multiple tests during and after the development process to ensure the stability and performance of the system. Make necessary optimizations based on the test results. Collect user feedback and make necessary modifications and optimizations based on the feedback. Deploy the system to a server or cloud and perform regular maintenance and updates.

Li et al. [9] designed a reliable data management system to store and manage student information, course information, grade information, etc. This system is an expandable teaching resource management system to support various forms of teaching resources, such as videos, images, documents, etc. It optimized the user interface design and designed a user-friendly interface to support the use of teachers, students, and administrators. Finally, a comprehensive security plan was designed. In terms of specific implementation, network programming languages such as Java, PHP, Python, and database systems such as MySQL and Oracle, as well as web servers such as Apache and Nginx, can be used to implement the system. Cloud computing technology can provide effective support for the sharing and optimization of teaching resources. Through cloud computing technology, teaching materials, courses, learning tools, and other resources can be stored in the cloud, and learners can access these resources through any connected device. This storage and access method can greatly improve the utilization and sharing of teaching resources, reduce educational costs, and improve educational efficiency, aiming to provide users with more accurate and personalized recommendation services. Traditional similarity calculation methods may not fully consider the complex relationship between users and items. Therefore, the algorithm adopts an improved similarity calculation method, such as cosine similarity, Jaccard similarity, etc., to better measure the similarity between users. Li et al. [10] used the forgetting curve model to calculate the user's interest weight, and then made recommendations based on the interest weight. By combining these two methods, the algorithm can more accurately predict users' interests and provide more personalized recommendations. In addition, the algorithm can be further optimized and improved according to actual situations, such as introducing more recommendation strategies, adjusting parameters, etc., to improve recommendation effectiveness. Ojokoh et al. [11] analyzed the characteristics of movies (such as content, style, actors, etc.) and the historical viewing behavior of users to provide a personalized movie recommendation system. In the international business environment, such a system can provide an effective market strategy to help film companies increase their sales and visibility. Based on movie characteristics and user behavior, the system can generate personalized movie recommendation lists. This may involve some complex machine learning models, such as collaborative filtering, content-based recommendation algorithms, etc. The courses should include the following modules: course management, resource management, student management, homework management, and communication management. The course management module can help students organize their learning according to the course. The homework management module can manage student assignments, while the communication management module can achieve online communication and feedback. The implementation of courses can be achieved through web application programming, using network programming languages such as Java, PHP, Python, and combining with database systems such as MySQL and Oracle, as well as web servers such as Apache and Nginx. With the support of cloud computing technology, various teaching resources can be stored in the cloud, and students can access these resources through the network to achieve anytime, anywhere learning. Pei and Wang's [12] use cloud computing technology to manage and schedule teaching resources, achieving optimal utilization of resources. This system analyzes students' historical learning behavior and interest preferences, combined with the similarity and forgetting curve of music, to recommend more personalized and suitable music resources for students' learning needs. Quan [13] classifies and recommends teaching resources through statistical analysis of their usage, making it easier for students to find the learning resources they need. Suitable algorithms can be selected, such as decision trees, support vector machines, random forests, etc., to train the model and make predictions by analyzing historical data. This module uses KNN algorithm to classify and cluster students, in order to provide personalized teaching suggestions for teachers. Suitable similarity measurement methods can be selected, such as Euclidean distance, cosine similarity, etc. By analyzing student features, similarity can be calculated and classified. Through the above development, a computer-aided English classroom teaching system based on machine learning prediction and artificial intelligence KNN algorithm can provide students with a more intelligent, personalized, and autonomous learning experience. At the same time, it can also provide teachers with more convenient and efficient teaching management tools, achieving the sharing and optimization of teaching resources.

The PZB model (Pearson Zhang Baker model) is an English testing model based on Item Response Theory (IRT). The model parameters reflect the students' English proficiency, the difficulty parameter represents the difficulty of the test question, and the guessing parameter describes the students' tendency to randomly guess. Through the PZB model, students' English proficiency can be analyzed, the difficulty and discrimination of test questions can be evaluated, and students' guessing behavior can be understood. The TAM model (Technology Acceptance Model) is a theoretical model used to study users' acceptance of information technology. Through the TAM model, you can understand students' acceptance of English evaluation and teaching equipment, and design and optimize the system based on students' feedback information. When designing an English evaluation and teaching equipment system, Tan et al. [14] combined the PZB model and TAM model to comprehensively consider factors such as students' English proficiency, difficulty of test questions, students' guessing behavior, and students' acceptance of the system. Understand students' acceptance of the system through the TAM model, and design and optimize the system based on this information to improve students' learning effectiveness and teachers' teaching efficiency. Sharing and interconnecting learning resources to improve learning efficiency. Wang [15] explored how to combine IoT technology with artificial intelligence algorithms to achieve personalized resource recommendation for intelligent English education. The intelligent English education platform based on the Internet of Things is an open platform that connects various learning resources. The platform can achieve the sharing and interoperability of learning resources, providing learners with rich learning materials. At the same time, the platform also integrates various artificial intelligence algorithms to analyze learners' learning behavior and interests, and recommend suitable resources for them. This corpus will contain parallel texts of the source and target languages, which can be used for teaching and translation research. Wang [16] chooses a suitable computer-aided translation software for analysis and research. Some free open source software such as Google Translate and Youdao Translation can be used to build such corpora. The next step is to collect corpus for English translation. This can be achieved in various ways, including downloading publicly available translation files from the internet or obtaining them from existing translation databases. After collecting corpus, it is necessary to align and proofread these texts to ensure accurate correspondence between the source and target languages. This can be achieved by using specialized tools such as Adobe Acrobat Pro. After aligning the corpus, annotations and annotations can be added to help learners better understand the translation process and techniques. Annotations can include information on vocabulary, grammar, cultural background, and other aspects.

Xu [17] will integrate the constructed corpus into the teaching environment, allowing students to freely access and use it both inside and outside the classroom. This can be achieved by developing specialized teaching platforms or using existing online learning management systems. By building a parallel corpus, students can learn vocabulary and grammar in real contexts, thereby improving their translation skills. Teachers can also use this corpus for classroom teaching and translation research. This method has broad application prospects in modern translation education. Including their English proficiency, learning preferences, learning styles, and historical learning behaviors. This can be achieved through the use of testing tools, questionnaire surveys, student evaluations, and other methods. Based on the collected data, establish a model for each student. This model should reflect students' English proficiency, learning preferences, and learning styles. This can be achieved through the use of data mining and machine learning algorithms. With the acceleration of globalization and informatization, the role of English in daily life and work is becoming increasingly important. For students and professionals, finding high-quality English online teaching resources is an important task. However, current online teaching platforms often have some problems, such as slow access speed, unfriendly interface, and untimely resource updates, which affect users' learning experience and usage effectiveness. Therefore, Zhang [18] analyzed that an English online teaching resource sharing platform based on mobile network technology can record students' learning behavior and performance. And provide personalized learning suggestions and resource recommendations based on students' characteristics and needs. This helps to achieve students' autonomy and personalized learning. The platform will continuously update and enrich teaching resources to meet the growing learning needs of users. In addition, the platform will also strengthen user interaction and community building, improve user engagement and stickiness. In short, the English online teaching resource sharing platform based on mobile network technology will provide convenient and efficient learning paths for students and professionals, effectively promoting the development of English education. Zhao and Guo [19] adopt a B/S architecture, with a web browser as the client, and the server side includes a database server, application server, and other designed intelligent auxiliary teaching modules. This module is the core of the system, including course management, course resources, online assignments, testing and evaluation, and other functions. Course management can help students organize their learning according to the course, and course resources provide various learning materials. Including videos, audio, documents, etc., online assignments allow students to complete and submit assignments within the specified time, while testing and evaluation tests and evaluates students' learning effectiveness. This module includes functions such as corpus construction, maintenance, and management. Corpus is the core database of the system, including student assignments, test papers, teacher resources, and other content. By analyzing and managing the corpus, the system can intelligently identify students' learning problems and provide personalized learning suggestions. Speech recognition can recognize students' oral expressions, natural language processing can analyze students' compositions, and machine learning can train models and optimize system performance based on students' learning behaviors. The goal of this system is to provide personalized book recommendation services by analyzing the borrowing history of readers and the association rules between books. Zhou [20] obtains readers' borrowing history data from the library's borrowing system and performs necessary preprocessing. The system uses an improved Apriori algorithm to mine association rules between books. This algorithm improves the efficiency of mining by reducing the number of candidate item sets. It can also discover association rules between different book categories, enriching the diversity of recommendation results.

3 DM AND KDD MODEL OF CALL SYSTEM

3.1 CAI System DM

In order to better meet the demands of online education users to choose their own courses, users' own characteristics need to be fully considered in the course recommendation algorithm. By collecting the data existing in the network, and then calculating its association, we can judge the intended resources and complete the user recommendation. Although this algorithm is simple, intuitive and easy to operate, it will ignore the individual characteristics of users when calculating the relevance of information, and there will be cases where expired data cannot be cancelled or shielded, thus affecting the accuracy of calculation and reducing the overall recommendation efficiency. This algorithm makes the best use of the timeliness characteristics of users, which makes the preference easier to calculate and obtain, but ignores the fact that the information itself is also timeliness, which leads to the situation that the two can not coexist easily and affects the accuracy of data recommendation. However, only relying on historical behavior to obtain interest, there are still shortcomings in extracting users' personal characteristics and accurately grasping users' interest preferences.

Individualized service recommendation can make individualized recommendations for users by analyzing their relevant information. Such recommendation results are more in line with users' personal preferences, thus saving users time. With the wide use of emerging media technology, the quantity of online education platform users and courses is increasing. However, users' rating data for courses is relatively small, and users' hidden preferences are easy to be missed in the case of insufficient data, and each user's own characteristics are also different. These factors will affect the satisfaction of the CAD system, so many course CAD systems begin to look for new solutions. By introducing user portraits to fully display the user characteristics, it can help the recommendation algorithm to understand the user characteristics. The resource integration stage of the CALL system proposed in this article is shown in Figure 1.

Interaction is unique to computers and multimedia computers, which makes multimedia computers an effective instructional method. In order to judge recommendable data and non-recommendable data more intuitively and accurately, it is needed to carry out the corresponding depth DM according to the recommendation requirements of different users, and then calculate the similarity between data information by using CF algorithm, so as to realize the accurate distinction of recommended data and ensure the recommendation efficiency and work quality of subsequent resource recommendation models. The system design fully considers students' learning habits and learning methods, and organically combines the characteristics of English vocabulary to find out the main problems in students' English learning.

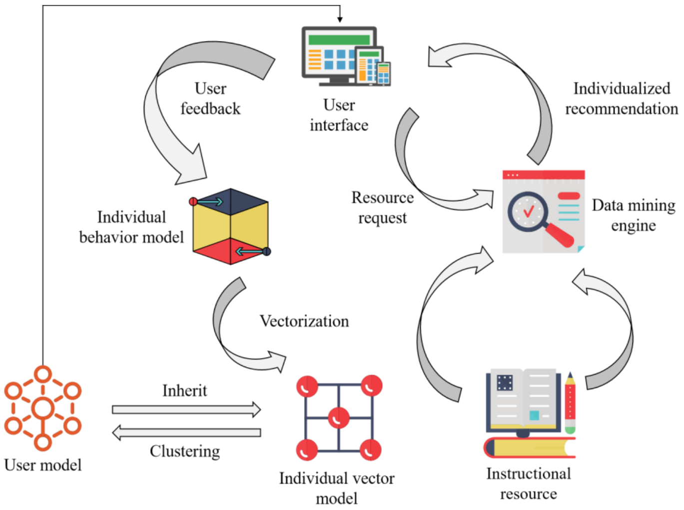

Figure 1: Resource integration process.

Students' scores on courses represent students' interests to a great extent, but a large quantity of students' evaluation data have not been effectively used. If students' evaluation data can be used to make individualized course recommendations according to each student's own hobbies, the blindness of students' course selection can be effectively solved. Through the online users' choice information of platform courses, the internal relationship of courses is explored, and finally the course sequence is recommended for the target users according to the collaborative information modeling of courses.

In order to measure the attribute weights more reasonably, the weights need to be normalized. After defining the normalized weight set, the weighted support definition of attribute set $B=\{b_1,b_2,\cdots,b_m\}$ is given:

$$w\sup(B)=\sum_{i=1}^{p} w_{b1}\times\frac{support(B)}{n} \tag{1}$$

If $wsup(X)$ is not less than the minimum support given by the user, B is called weighted frequent attribute set.

The weighted support degree of weighted Boolean association rule $B\Rightarrow B^{'}$ is defined as:

$$w(\sup)=\sum_{i=1}^{m+n} w_{b1}\times\frac{support(B\cup B^{'})}{n} \tag{2}$$

The weighted confidence of weighted Boolean association rule $B\Rightarrow B^{'}$ is defined as:

$$w(conf)=\frac{support(B\cup B^{'})}{support(B)} \tag{3}$$

The nature of weighted Boolean association rules: the support strategy of hierarchical association rules is adopted to specify different support thresholds for different conceptual levels. The support thresholds of the upper level are relatively large, while those of the lower level are relatively small.

Data integration is based on the deeply excavated recommended data, and the recommended data is standardized, revised and matched. After verification, the successfully matched data is imported into the integrated database. The data that failed to match are marked and then imported into the integration database. According to different management units, the processing modules with different functions can gradually adjust during the integration process, or the data integration process can be started regularly to clean and sort out the data during the operation of the recommended model.

In order to effectively solve the problems existing in the traditional similarity measurement method when students' grading data for courses are extremely sparse, the similarity between courses is first calculated according to the characteristics of courses themselves. Then the courses are clustered by clustering method, and the nearest neighbors of the courses are found in the same cluster. According to the students' scores of the nearest neighbor courses, the score of the target courses is predicted, and a course scoring matrix without missing is constructed.

Define the inter-cluster distance L as the sum of distances from all cluster centers to all data centers. The calculation formula of L is shown in (4):

$$L=\sum_{i=1}^{K}\left\|c_i-m\right\| \tag{4}$$

Where c_i is the result cluster center, m is all user data centers, and k is the quantity of clusters.

The intra-cluster distance D is defined as the sum of the total distances from all points in each cluster in all the result clusters to the center of the cluster, x represents a point in the result cluster. The calculation formula of D is shown in (5):

$$D=\sum_{i=1}^{K}\sum_{x\in C}\left\|x-c_i\right\| \tag{5}$$

3.2 KDD Model Based on Improved CF Algorithm

In order to improve the efficiency and accuracy of information recommendation, it is needed to build information data sets of English databases in a large quantity of environments, and complete the collection and reception of information by passing them one by one. Figure 2 shows the DM and KDD model structure based on CF algorithm.

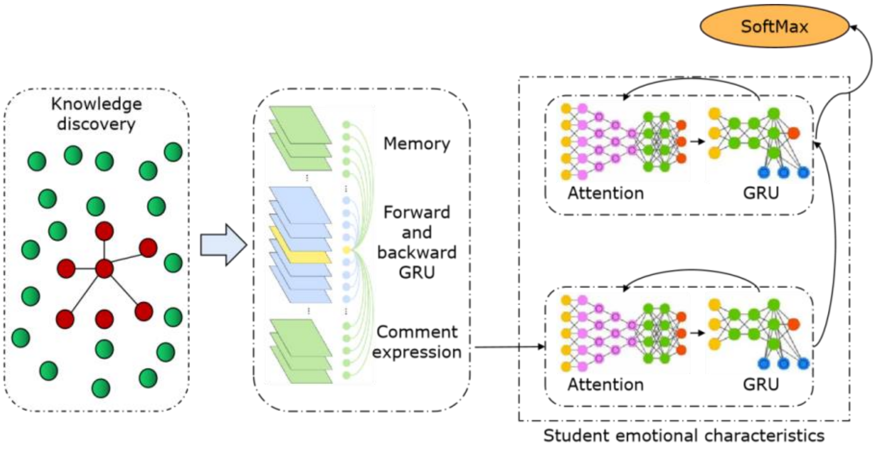

Figure 2: DM and KDD model.

The memory module uses a bidirectional Gate Recurrent Unit (GRU) network to model the context semantic information, so as to obtain a more accurate representation of context features; And use attention mechanism and GRU unit to selectively enhance or forget prior knowledge. Update the prior knowledge by superimposing multiple computing layers; The output module takes the emotion expression of the output of the last computing layer as the input of the SoftMax classifier, and realizes preference classification.

In order to classify students' preferences more accurately, it is needed to make as many comments as possible and extract more accurate emotional features by using context dependence. The simplified GRU model of can be expressed as:

$$h_t = GRU(x_t, h_{t-1}) \tag{6}$$

When the context $X = \{x_1, x_2, \cdots, x_n\}$ of the comment is input, the hidden layer state $\vec{h}_t$ is output to GRU:

$$\vec{h}_t = GRU(x_t, \vec{h}_{t-1}) \tag{7}$$

Through LDA (Latent Dirichlet Allocation) topic model output, the result is n topics, and each topic has m words closely related to it. Definition w_{ij} stands for the j th word related to the i th topic. Expressed in matrix form as:

$$\begin{bmatrix} w_{11} & w_{12} & \cdots & w_{1m} \\ w_{21} & w_{22} & \cdots & w_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ w_{n1} & w_{n2} & \cdots & w_{nm} \end{bmatrix} \tag{8}$$

After paying attention to memory for many times and extracting information, the output vector y_i^t of the last calculation layer is the final emotion feature y_o, which can be interpreted as a conditional probability, and it is input into a SoftMax classifier for preference classification:

$$p_c = \frac{\exp(W_s \cdot y_o + b_s)}{\sum_{c=1}^{c} \exp(W_c \cdot y_o + b_c)} \tag{9}$$

Where, C is the quantity of categories, and p_c is the probability of being predicted as category c.

According to statistical analysis, there is a strong correlation between adjacent courses. First, we should predict its corresponding context information for each course, take the course as the target vector, and select the quantity of choices on both sides of the current course by using the window size. Calculate the window-based course probability distribution, determine the co-occurrence probability of other courses and the current course, and take the higher probability as the next recommendation result.

The design of this system mainly takes into account students' study habits, learning methods and other factors, and finds out the main problems existing in students' English learning by combining the basic characteristics of computer English vocabulary. According to the problems, a new learning model is established to improve students' initiative in the stage of computer English learning. After CAI introduced ELT, it changed the traditional teaching mode, formed a new teaching concept and deepened quality education. Finally, the learning effect is verified and tested according to the classification of computer terms and the characteristics of computer English. After learning, users can maintain the question bank, count the scores and analyze the learning effect.

4 RESULT ANALYSIS AND DISCUSSION

By studying the relationship between users and courses, the course recommendation algorithm finds user preferences according to users' own characteristics, and further gives the course recommendation results. To achieve the expected purpose of the experiment, a quantity of large and well-known online learning platforms in China were comprehensively referenced in the data collection stage. The online course selection data set of an open university is used for the experiment. This data set contains information of learning users, course-related information and information table of elective courses for users. The data capture method used in this study is the Scrapy framework in Python computer programming. Scrapy Engine is the core structure of the whole crawler, and its main function is to control the trend and progress of the whole process and hand over the progress process.

The change of the Area Under Curve (AUC) value of this method with the sample size is tested, and the other two classification methods based on emotion ontology and word vector model are compared. The ratio set is cross-verified 10 times according to the K-Fold Cross-Validation method, and the final results are shown in Figures 3 and 4.

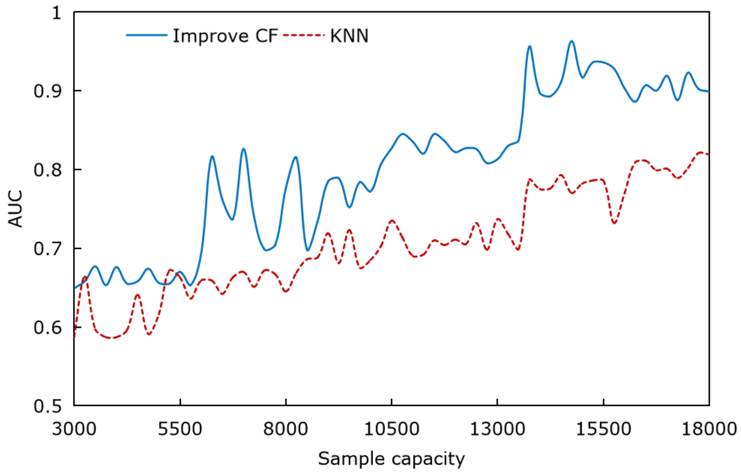

Figure 3: Classification effect of different sample sizes (20 dimensions).

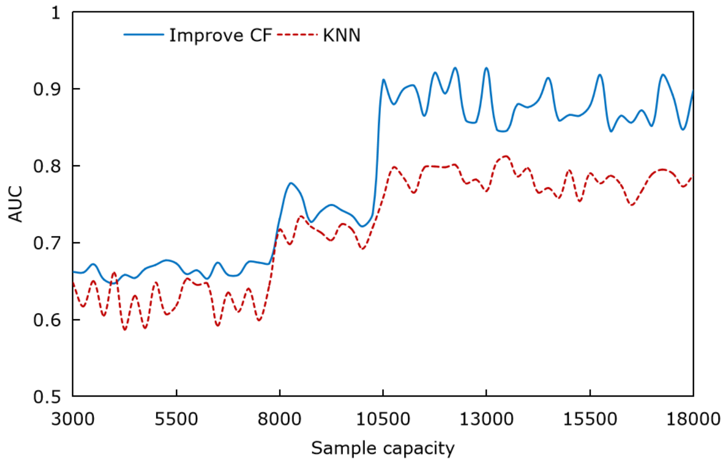

Figure 4: Classification effect of different sample sizes (80 dimensions).

As can be seen from the above figure, when the quantity of samples is greater than 10000, the AUC values of the three classification methods with different dimensions are all greater than 0.8 and tend to be stable, and the classification effects of the three classification methods reach a good level at this time. Curriculum collaborative algorithm can obtain the potential relationship between courses according to the historical behavior of users, which is helpful to improve the recommendation accuracy.

The accuracy of the recommendation results is compared and analyzed. Experiments are carried out on this algorithm and the traditional recommendation algorithm respectively, and the effectiveness of the improved KDD model is verified by comparing the prediction results. Figure 5 shows the comparison of the accuracy of algorithm recommendation results.

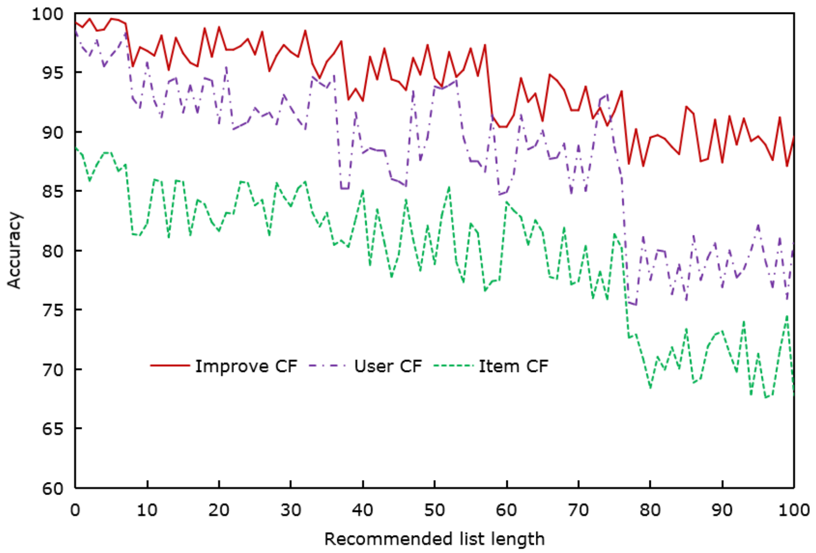

Figure 5: Comparison of algorithm recommendation accuracy.

The length of the recommendation list, mainly because the longer the recommendation list is, the more courses that users are not interested in will be introduced, which leads to the decrease of the accuracy of the algorithm. The algorithm in this article can maintain the highest accuracy under different recommended list lengths, especially when the recommended list length is 20, and its accuracy is obviously higher than the other two algorithms, which shows that the recommended results of this algorithm have higher accuracy.

Figure 6 shows the change of Recall index with the nearest neighbor K between this algorithm and other algorithms with reference to time weight factor. Considering the dynamic changes of user interest and time decay, we can see that the recommendation algorithm proposed in this article is not significant when the quantity of nearest neighbors is not very large. The accuracy index is higher than other comparison algorithms.

Under different quantity of neighbors, the course recommendation effect obtained by this recommendation algorithm is better than KNN recommendation algorithm. With the same quantity of neighbors, when the quantity of neighbors is 10, the Precision and Recall of this algorithm are 0.274 and 0.189 higher than KNN algorithm, respectively. This also shows that the course recommendation algorithm proposed in this article has achieved more remarkable results through the improvement of accurately grasping students' interest preferences. On the other hand, the numerical quality of each index of the traditional recommendation method is low, mainly because it does not integrate the original data before implementing the recommendation, resulting in sparse and uneven information, increasing the difficulty of recommendation, reducing the recommendation efficiency and affecting the overall effect.

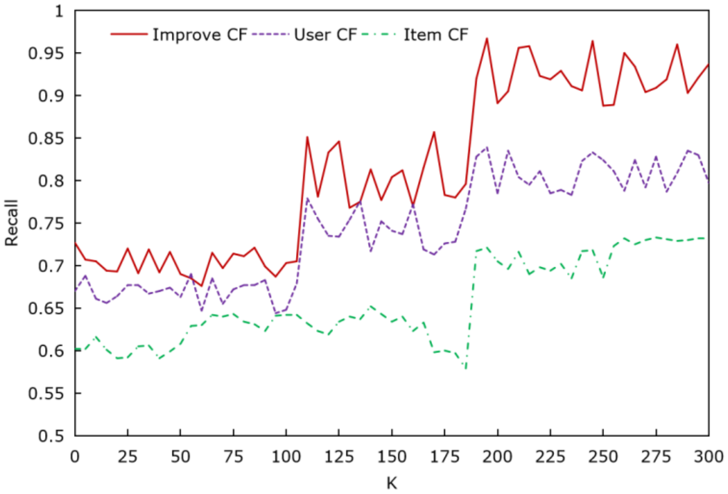

Figure 6: Recall index of the algorithm under different quantity of neighboring users.

When there is less information about a new user in the system, you can rely more on the project-based information prediction score; When new projects are added to the system or there are few project-related scores, we can rely more on the method based on user characteristics to calculate the score value. Figure 7 shows students' subjective scoring results of traditional ELT course acquisition method and recommendation based on improved CF algorithm.

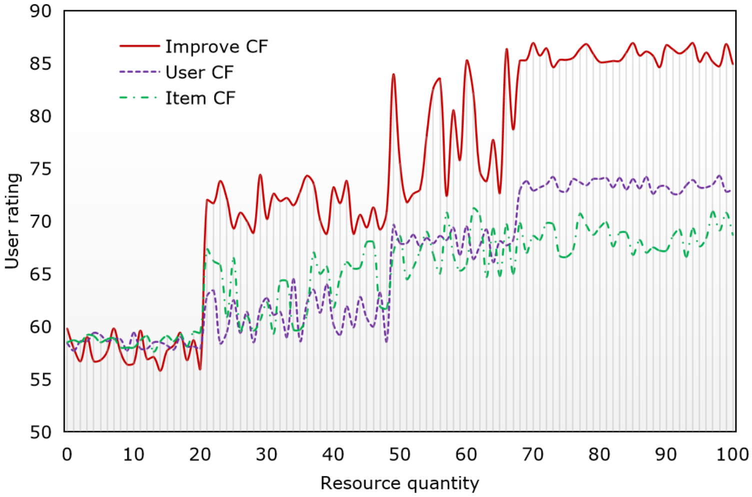

Figure 7: Students' subjective rating.

Most students say that the improved recommendation method accurately and tap their potential interests, thus providing more targeted course content. The algorithm in this article performs well in all indicators, and has strong self-adjustment ability, which can effectively improve the impact caused by information sparsity, adjust and fit the resource data with high information sensitive parameters and poor timeliness. This is of positive significance for the school to promote the construction of CAI system.

5 CONCLUSIONS

The increase of course content brings more choices for students to choose courses, but too much learning content also makes students' course selection inevitably blind. At present, the extensive application of teaching management system has accumulated a large quantity of instructional practice data, but the information implied in these data has not been fully exploited. The course CAD system realizes the recommendation function to meet the demands and interests of users by mining and analyzing the relevant information of users. This article presents a DM and KDD model of CALL system based on improved CF algorithm. Through the steps of data collection and deep mining, this method effectively improves the problem of resource sparsity caused by original data and improves the accuracy of recommendation algorithm. This course recommendation algorithm not only improves the sparsity of data to some extent, but also improves the dynamic change of users' interest preferences, which is conducive to improving the recommendation effect. By comparing the ways of users' feedback scoring on data, similar data can be effectively distinguished and information recommendation in English database can be completed efficiently. Experiments show that the improved algorithm proposed in this article has achieved good results. The next research focus is to further improve the scalability and ensure the real-time performance of the recommendation algorithm.

6 ACKNOWLEDGEMENT

This work was supported by the Teaching Reform Research Project of Shaanxi Fashion Engineering University in 2022: The Curriculum Ideology and Politics Teaching Team of College English (2022JXTD-06); This work was supported by the Teaching Management Project of Shaanxi Fashion Engineering University in 2023: Strengthen the Research on the Design and Management of the Teaching Link of "Three Classes" Education under the Guidance of Ideology and Politics(2023JGZX024).

Furong Shi, https://orcid.org/0000-0002-8172-2640
Mengmeng Zhang, https://orcid.org/0000-0002-1368-4005

REFERENCES

[1] Alghamdi, A.-A.; Alanezi, M.-A.; Khan, F.: Design and implementation of a computer aided intelligent examination system, International Journal of Emerging Technologies in Learning (iJET), 15(1), 2020, 30-44. https://doi.org/10.3991/ijet.v15i01.11102

[2] Bi, S.: Intelligent system for English translation using automated knowledge base, Journal of Intelligent & Fuzzy Systems, 39(4), 2020, 5057-5066. https://doi.org/10.3233/JIFS-179991

[3] Chen, H.; Huang, J.: Research and application of the interactive English online teaching system based on the internet of things, Scientific Programming, 2021(1), 2021, 1-10. https://doi.org/10.1155/2021/3636533

[4] Chen, X.; Deng, H.: Research on personalized recommendation methods for online video learning resources, Applied Sciences, 11(2), 2021, 804. https://doi.org/10.3390/app11020804

[5] Fan, L.: The method of interest text recommendation in English education based on data mining, International Journal of Continuing Engineering Education and Life Long Learning, 32(3), 2022, 374-388. https://doi.org/10.1504/IJCEELL.2022.124031

[6] Gong, W.: An innovative English teaching system based on computer aided technology and corpus management, International Journal of Emerging Technologies in Learning (iJET), 14(14), 2019, 69. https://doi.org/10.3991/ijet.v14i14.10817

[7] Huang, Y.; Zhu, J.: A personalized English learning material recommendation system based on knowledge graph, International Journal of Emerging Technologies in Learning (Online), 16(11), 2021, 160. https://doi.org/10.3991/IJET.V16I11.23317

[8] Jing, D.; Jiang, X.: Optimization of computer-aided English teaching system realized by vb software, Computer-Aided Design and Applications, 19(S1), 2021, 139-150. https://doi.org/10.14733/cadaps.2022.S1.139-150

[9] Li, H.; Zhang, H.; Zhao, Y.: Design of computer-aided teaching network management system for college physical education, Computer-Aided Design and Applications, 18(S4), 2021, 152-162. https://doi.org/10.14733/cadaps.2021.S4.152-162

[10] Li, T.; Jin, L.; Wu, Z.; Chen, Y.: Combined recommendation algorithm based on improved similarity and forgetting curve, Information, 10(4), 2019, 130. https://doi.org/10.3390/info10040130

[11] Ojokoh, B.-A.; Aboluje, O.-O.; Igbe, T.: A collaborative content-based movie recommender system, International Journal of Business Intelligence and Data Mining, 17(3), 2020, 298-320. https://doi.org/10.1504/IJBIDM.2020.109293

[12] Pei, Z.; Wang, Y.: Analysis of computer aided teaching management system for music appreciation course based on network resources, Computer-Aided Design and Applications, 19(S1), 2021, 1-11. https://doi.org/10.14733/cadaps.2022.S1.1-11

[13] Quan, Y.: Development of computer aided classroom teaching system based on machine learning prediction and artificial intelligence KNN algorithm, Journal of Intelligent and Fuzzy Systems, 39(6), 2020, 1-12. https://doi.org/10.3233/JIFS-179959

[14] Tan, P.-J.-B.; Hsu, M.-H.: Designing a system for English evaluation and teaching devices: a pzb and tam model analysis, Eurasia Journal of Mathematics Science and Technology Education, 14(6), 2018, 2107-2119. https://doi.org/10.29333/ejmste/86467

[15] Wang, F.: IoT for smart English education: AI-based personalized learning resource recommendation algorithm, International Journal of Computer Applications in Technology, 71(3), 2023, 200-207. https://doi.org/10.1504/IJCAT.2023.132093

[16] Wang, X.: Building a parallel corpus for English translation teaching based on computer-aided translation software, Computer-Aided Design and Applications, 18(S4), 2021, 175-185. https://doi.org/10.14733/cadaps.2021.S4.175-185

[17] Xu, Y.: Computer-aided design of personalized recommendation in teaching system, computer-aided design and applications, 17(S1), 2019, 44-56. https://doi.org/10.14733/cadaps.2020.S1.44-56

[18] Zhang, Y.: Design of an English web-based teaching resource sharing platform based on mobile web technology, International Journal of Advanced Computer Science and Applications, 14(2), 2023, 572-580. https://doi.org/10.14569/IJACSA.2023.0140268

[19] Zhao, H.; Guo, L.: Design of intelligent computer aided network teaching system based on web, Computer-Aided Design and Applications, 19(S1), 2021, 12-23. https://doi.org/10.14733/cadaps.2022.S1.12-23

[20] Zhou, Y.: Design and implementation of book recommendation management system based on improved Apriori algorithm, Intelligent Information Management, 12(3), 2020, 75-87. https://doi.org/10.4236/iim.2020.123006