Research on English Vocabulary Learning Platform Based on Personalized Recommendation

Hongyan Cai*

* Corresponding author: 9903205@qq.com

Shandong Agriculture and Engineering University, Jinan 250100, China

Abstract— In the mobile learning environment, it is necessary to consider how to mine learners' personalized vocabulary needs and recommend appropriate vocabulary resources according to the characteristics of resources and learners. This is of practical significance for improving learners' vocabulary learning efficiency. Based on the classification time perception method and vector space model representation, this paper constructs the learner dynamic interest model. This model can represent learners' personalized features more accurately and improve the accuracy of word recommendation. Combined with the collaborative filtering recommendation based on learners' dynamic interests, it can recommend vocabulary resources for learners. Finally, through the questionnaire survey, the effectiveness of the system is tested from the aspect of the learning effect of the system. The data show that the system can improve the user's learning effect to a certain extent.

Keywords: personalized recommendation; spatial model; collaborative filtering; lexical recommendation; dynamic model

1 INTRODUCTION

Vocabulary is the foundation of a language, and mastering a language begins with learning vocabulary. Especially in English, the amount of vocabulary a learner has mastered and the level of proficiency affects his or her English proficiency [1]. In the process of the National College English Proficiency Test, learners need to master a certain amount of vocabulary. Therefore, as the foundation and key part of English learning, learners should pay attention to the learning of English vocabulary [2]. Currently, there are many vocabulary learning software. However, it takes a lot of time and effort to choose the most suitable software for learners from these distinctive software, which affects learners' motivation and incentive in memorizing vocabulary. Moreover, the current vocabulary learning system has a fixed learning mode, which makes learners easily lose interest and fails to stimulate learners' interest in learning [3]. Therefore, this study designs an English vocabulary learning recommendation system, which focuses on college English vocabulary, combines personalized recommendation technology and the guidance of related theories, conducts in-depth analysis of learners' personalized needs, and provides personalized vocabulary recommendations for different types of learners. The system improves learners' vocabulary learning efficiency and solves the defects of existing English vocabulary learning software that does not meet learners' personalized needs [4].

ITEI 2022, December 23-25, Harbin, People's Republic of China

DOI 10.4108/eai.23-12-2022.2329214

2 PERSONALIZED RECOMMENDATION PRINCIPLE AND RELATED RECOMMENDATION ALGORITHM ANALYSIS

2.1 Principle of personalized recommendation

The system of personalized recommendation generally contains three parts: user model, content model and recommendation engine [5]. The user model is a model based on information about the user. In this model the user's data information, such as user behavior, interests, etc., is transformed into computer-understandable data by certain algorithms and stored in the server.

Content models are for the information retrieved by users. Content models classify and model existing content according to their respective characteristics, with the aim of recommending content to target users more precisely [6]. The accuracy of the content recommended to users is directly proportional to the content model content, so it is necessary to build a large number of content models to improve the accuracy of content.

Recommendation engine is the key to personalized recommendation. It is the algorithmic bridge between user model and content model. First, it obtains users' data through big data, and then matches different algorithms to target users of different programs to precisely deliver the content they need [7].

2.2 Related recommendation algorithms

Currently, the common recommendation algorithms are: collaborative filtering recommendation algorithm, content-based recommendation algorithm and hybrid recommendation algorithm [8]. This paper is mainly based on the implementation of collaborative filtering algorithm, so this section focuses on the analysis of collaborative filtering algorithm.

Collaborative filtering recommendation algorithm is the most widely used personalized recommendation technique [9]. Its main idea is to calculate the similarity between users or items based on the current historical rating information, find similar groups of users or items based on the similarity, and get the recommendation list by the rating of similar groups. The collaborative filtering recommendation algorithm calculates users' predicted ratings of items based on their historical ratings, and recommends items with high predicted ratings to users as a candidate set.

The key to collaborative filtering is to calculate the similarity between objects and the similarity of users. The similarity formula is used to determine the preference of the selected users for this information. The general process of collaborative filtering is shown in Figure 1 below.

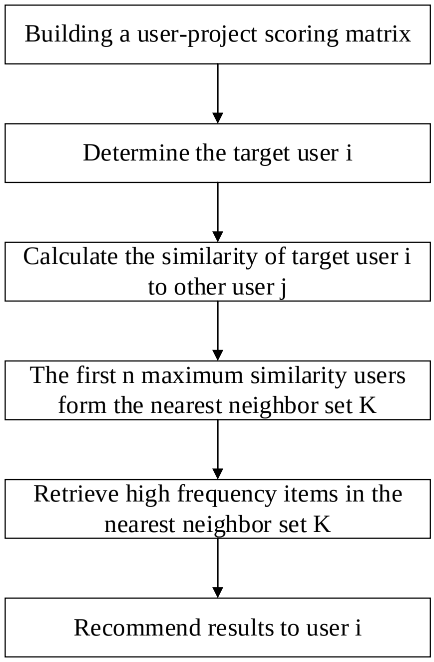

Figure 1. Collaborative filtering recommendation process

a) Data representation: First, a matrix of user ratings of items is constructed, where M denotes all users and N denotes all items. Then, a scoring matrix of M users on N items can be formed between users and items, and R_{mn} represents the scoring of user M on item N. The system in this paper is to form a user-item scoring matrix for the test questions. The resulting user-item scoring matrix is shown in Formula 1 below.

$$R = \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1N} \\ r_{21} & r_{22} & \cdots & r_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ r_{M1} & r_{M2} & \cdots & r_{MN} \end{bmatrix} \tag{1}$$

b) Similarity calculation: The similarity between target users is calculated by the similarity calculation formula, and then the similarity is corrected according to the modified cosine similarity formula. The Pearson similarity algorithm is used to solve the problem when users have similar preferences for items but have large differences in ratings.

In this paper, we calculate the learner-learner similarity $sim(a, b)$ in the learner-vocabulary category scoring matrix M based on the cosine similarity formula, and select the top N nearest neighbor user sets $S(a, N)$ based on the magnitude of similarity.

$$sim(a, b) = \frac{\sum_i r_{a,i} r_{b,i}}{\sum_i {r_{a,i}}^2 \sum_i {r_{b,i}}^2} \tag{2}$$

c) Formation of neighbor set: Calculate the similarity between target user i and user j, and form the nearest neighbor set K by the top N users with the maximum similarity.

d) Generate recommendation dataset: Retrieve the high frequency items in the nearest neighbor set K, select the top items with the most value to form the recommendation dataset, and then recommend them to the target users.

3 LEARNER DYNAMIC LEARNING MODEL CONSTRUCTION

Learners who are interested in certain vocabularies will concentrate on learning those vocabularies and can improve their learning efficiency [10]. Therefore, analyzing learners' learning interests is important for constructing learner models. Since learners' interests change constantly, it is important to construct a model of learners' dynamic interests.

In this paper, a dynamic interest model for learners based on vocabulary categories is designed based on a categorical time-aware approach and a vector space model representation. The learner's behavioral data is used to dynamically analyze the learners' preferences for vocabulary resources, and the model is constructed based on the dynamic interest model of vocabulary categories, and the model construction steps are shown below.

1) Firstly, by the categorical time perception method, the time is divided into several time windows according to the temporal behavior characteristics of learners accessing the system, and defined as $T = \{T_1, T_2, \ldots, T_n\}$, where Tn refers to the nth time window. The learner's interest is expressed as the vocabulary category he/she learns, and there are thirteen vocabulary categories, and the set of vocabulary categories is defined as $C = \{C_1,\ C_2, \ldots, C_{13}\}$. In this paper, the specific time of the time window is set to one day. The learners' interest in vocabulary learning is updated daily.

2) The learner-vocabulary category preference weights are calculated for the set of learner preference data according to the division of time windows. Since there are thirteen vocabulary categories and learners may be interested in multiple vocabulary categories, the learner's preference weights for each vocabulary category need to be calculated.

3) To construct an interest model for the learner's behavior under the current time window Tn, the learner's interest in learning under each time window is expressed as the category of vocabulary learnt by the learner under that time window. First, the learner's preference weight for a certain vocabulary category under the current time window is calculated by the following formula.

$$P_n = \frac{T_{c_n}}{T} \tag{3}$$

In which, P_n indicates the learners' preference for vocabulary category c_n, T_{c_n} indicates the total time that the learners visited the vocabulary containing vocabulary category c_n, and T denotes the total time that the learners learned n vocabulary categories under this time window.

4) The learner's preference weights for the thirteen vocabulary categories under the current time window are calculated, and finally the learner's dynamic interest model based on the vocabulary categories under the current time window is obtained as follows.

$$XDXM = P_1 + P_2 + \cdots + P_{13} \tag{4}$$

At each time window T_n, the learner's learning interest at that time window is represented by an n-dimensional vector through a vector space model representation, where XDXM represents a function of the learner's preference weights for each vocabulary category at the current time window.

4 COLLABORATIVE FILTERING RECOMMENDATION AND EXPERIMENT BASED ON LEARNER DYNAMIC INTEREST MODEL

4.1 Collaborative filtering recommendation algorithm design based on learner dynamic interest model

The overall recommendation strategy of the English vocabulary learning recommendation system built in this study is as follows: when a learner enters the system, the system first detects whether the existing database contains the vocabulary learning behavior data of the current learner. If there is no vocabulary learning behavior data of the current learner or too little vocabulary learning behavior data in the system, the system cannot determine the vocabulary learning needs of the learner, and the system will recommend vocabulary resources to the learner based on the initial model of the learner. If the current learner's vocabulary learning behavior data in the system can support the system's analysis of the learner's vocabulary learning needs, the system will initially analyze the learner's vocabulary level and learning interests based on these data, and recommend subsequent vocabulary resources for the learner by combining the learner's vocabulary level and dynamic interest model. The system first finds users with similar vocabulary ability level to the learners through dichotomous K-means clustering algorithm to reduce the complexity of similarity calculation, and then further finds users with similar learning interests in the set of users with similar vocabulary ability level to the learners through the recommendation strategy of collaborative filtering based on learners' dynamic interests based on waterfall hybrid technology, and counts the frequency of all vocabulary resources among similar users. The frequency of all vocabulary resources among similar users is counted, and the vocabulary learning resources are ranked according to the frequency, and finally the vocabulary resources are recommended to learners according to the ranking. The recommendation process is shown in Figure 2.

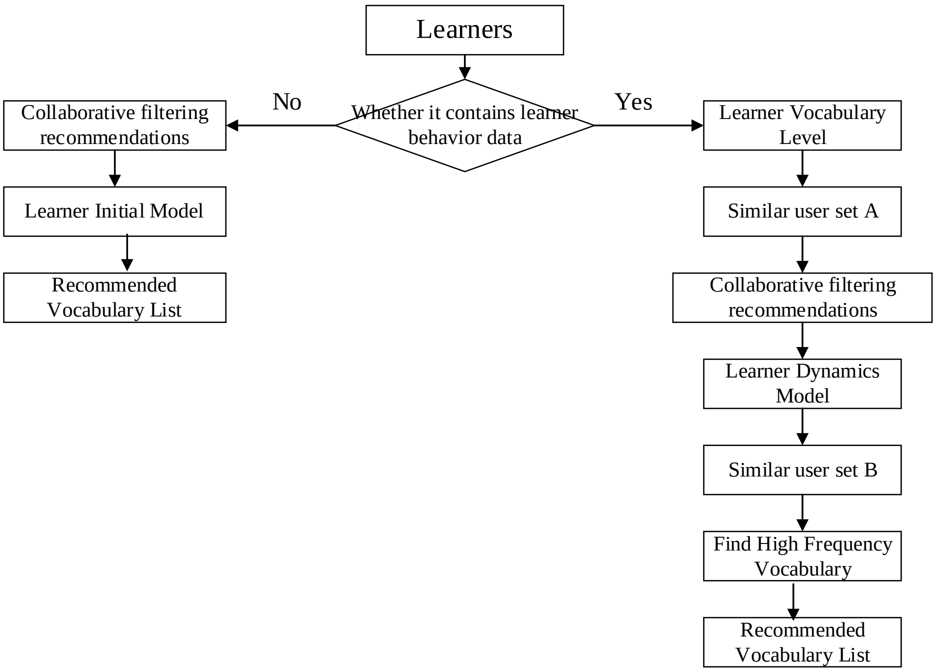

Figure 2. The overall strategy of vocabulary resource recommendation

Through the recommendation strategy of collaborative filtering based on the dynamic interests of learners, we find the set B of users with similar learning interests to learners, and finally count the frequency of all vocabulary resources in the similar user set B. The vocabulary learning resources are ranked according to their frequency, and the high frequency vocabulary resources are recommended to learners. The vocabulary learning resources are ranked according to their frequency, and the high-frequency vocabulary resources are recommended to learners.

The construction of the scoring matrix and the calculation of similarity have been analyzed above and will not be repeated here.

The frequency of each vocabulary learning resource in the vocabulary learning records of all similar learners, if the most frequent vocabulary resource is k2, followed by k1 and k3, is calculated, and the vocabulary learning resources are ranked according to the frequency of all similar learners, and finally the vocabulary resources are recommended to learners according to the ranking. As shown in Figure 3.

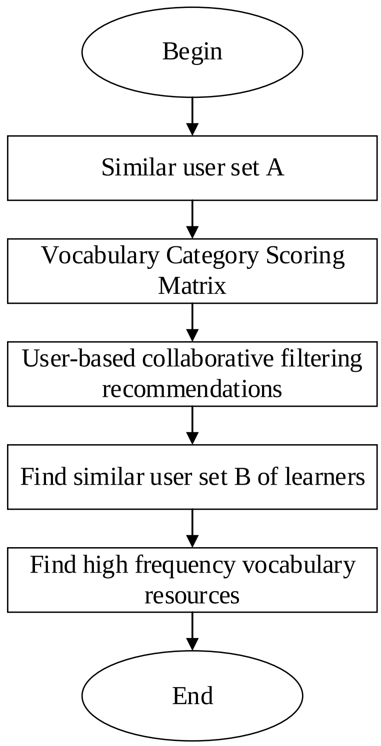

Figure 3. Collaborative filtering recommendation based on learner dynamic interest model

4.2 Experimental verification

Based on the above algorithm, an English learning platform was designed and implemented, and the learning effect was analyzed through a survey questionnaire. The analysis of whether the system can improve the efficiency of learners in learning English vocabulary at level 4 is shown in Figure 4. 57.69% of learners strongly agree that the system can improve the efficiency of learning English vocabulary, 23.08% of learners think the system can basically improve the efficiency of learning English vocabulary, a few learners think the system has little or no effect on learning English vocabulary, and overall learners agree that the system is useful for memorizing vocabulary.

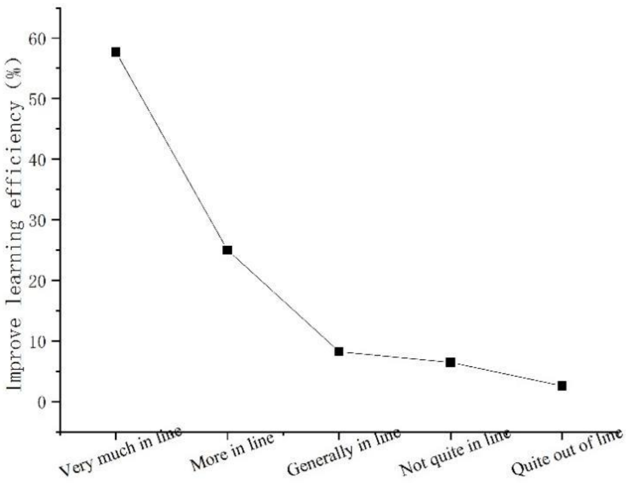

Figure 4. Learning Effect Analysis Chart

5 CONCLUSION

Based on the characteristics of college English vocabulary and learner behavior data, this paper constructs a multidimensional learner model, combines personalized recommendation technology, designs and implements an English vocabulary learning recommendation platform, and verifies the effectiveness of the platform. Providing learners with personalized vocabulary recommendations can improve learners' learning efficiency and solve the shortcomings of existing English vocabulary learning software that does not meet learners' personalized needs. In the future, we will continue to improve the system. More learners can improve their learning efficiency with the help of this system.

REFERENCES

[1] Ru S. Y., Deng J. D., Zhang Z. W., et al. (2020) Research on the application of English word learning system based on deep learning. Computer Programming Skills and Maintenance,02:133-134+147.

[2] Zhong Y. M. (2021) Algorithm design of personalized course recommendation model for online learning system. Computer Knowledge and Technology, 17(25):130-131+143.

[3] Lei K. W., Wang X. H. (2022) Design of personalized recommendation system for scenic spots based on WeChat public platform. Information Technology, 01:56-61.

[4] Liu Y. (2022) An online English learning system based on students' personalized needs. Automation Technology and Applications,41(09):161-164.

[5] Liu J., Zhang J. (2022) Design of personalized recommendation method based on data mining for short videos of painting and calligraphy. Journal of Changzhou Institute of Technology,35(02):30-34.

[6] Qin Q. H. (2022) Research on personalized recommendation system based on collaborative filtering algorithm. Science and Technology Information,20(10):4-6.

[7] Liu X. W. (2022) Research and application of personalized recommendation in e-commerce based on big data era. Modern Business, 15:23-25.
[8] Qin P., Jia H. J., Huo X. Y., et al. (2022) A user personalized POI recommendation method incorporating big data mining. Computer Simulation, 39(06):355-358+385.
[9] Yuan C. X. (2022) Research on the application of collaborative filtering algorithm in personalized recommendation in university libraries. Wireless Internet Technology,19(14):111-113.
[10] Jia Z. H., Bi H. J., Bai T. Y. (2022) Personalized recommendation algorithm for large open online course resources. Automation Technology and Applications, 41(08):146-149.