Personalized Recommendation Method for English Teaching Resources Based on Artificial Intelligence Technology

Meiyun Gao, Jun Xing*, Chengbo Yin, Linlinm Dai

College of Humanities, Hunan Broadcast and Television University, No.168 Qingyuan Road, Changsha, 410004, China

*862203480@qq.com

Abstract. Traditional recommendation methods decompose keywords and key sentences less frequently, resulting in low accuracy of recommended content. Therefore, a personalized recommendation method for English teaching resources based on artificial intelligence technology is proposed. Web mining technology is used to collect user personalized data and mine teaching resource rules and patterns that can represent user characteristics. Artificial intelligence technology is used to simulate the intelligent behavior of users searching for resources, to obtain words with a higher frequency. The words with higher relative contribution are selected as keywords, thereby calculating the similarity between keywords and user characteristics, then the teaching resources with high similarity are taken as the recommendation target, and the recommended format is selected to implement resource recommendation. Experimental results show that compared with traditional methods, this method improves the accuracy of recommended content and enables teaching resources to better meet the individual needs of users.

Keywords: Artificial intelligence; English teaching; Teaching resources; Personalized recommendation

1. Introduction

At this stage, the utilization rate of online teaching resources by teachers and students has increased, but due to the rapid increase in the number of teaching resources, it is increasingly difficult to find the teaching resources that are really needed in the process of searching for resources. Therefore, the personalized recommendation of teaching resources is of great significance [1]. The research on personalized recommendation of foreign online teaching mainly conducts in-depth analysis of users, and builds a comprehensive recommendation strategy based on users' demand for recommendation services. It includes a variety of recommendation service technologies, configuring the learning scenarios required by users, using hybrid filtering recommendation technology to expand the scope of application of learning scenarios, and filtering the content of teaching resources. And according to the user's recent navigation history and interactive behavior, the teaching resources are integrated to realize the self-adaptation of the user's network environment and complete the teaching resource recommendation. In terms of domestic teaching resource recommendation, the recommendation model and personalized recommendation technology are used to mine user historical data, collect and analyze learning behavior information. The interest information is obtained based on personal data, so as to determine the user's learning motivation and cognitive level, share and evaluate teaching resources,
count the differences in data information, and use association rules to build interest models, and recommend the generated teaching resources to users[2]. However, traditional personalized recommendation methods still have limitations, resulting in data sparseness and other problems. Therefore, a personalized recommendation method for English teaching resources based on artificial intelligence technology is proposed. Artificial intelligence technology can simulate intelligent behavior, analyze the thinking process of users' learning, and start from the user's logical thinking to provide reference for the recommendation of teaching resources.

2. Source personalized recommendation method design

2.1. Mining users’ personalized data
Web mining technology is used to mine users’ personalized data and describe the data in detail. According to the discovery rules of Web data mining technology, in-depth analysis of the user's Web log and page content is carried out to dig out interrelated information, obtain the information data that users are interested in, and divide the data information into two major categories: structure and content. Among them, the data structure is based on the overall organization of teaching resources. According to the relationship between the site and the page, a series of rules for personalized data are inferred to obtain the user’s knowledge model. For the data content, the relevant information in the access log file is selected to analyze the user's behavior pattern of searching for teaching resources, and extract the information and semantics of personalized data from it, thereby redefining the structure of English teaching resources[3]. After collecting the user's personalized information data, a user description file is created. The specific file structure is shown in the following table:

| Field         | Description          | Field           | Description                          |
|---------------|----------------------|-----------------|--------------------------------------|
| Admin         | User login name      | RemostHost      | How to access the page               |
| Date          | Length of time to visit the page | Method          | Communication protocol used          |
| AuthUse       | Number of bytes transferred | URL             | URL of the page                      |
| Protocol      | Remote machine name  | Bytes           | Server status                        |
| Status        | Authorized users of the server | Agent          | Lead page                            |

Tab.1 Personalized data description content

In the process of personalized data collection, the user's client and proxy server are selected to collect data from the server. Then the personalized data is preprocessed, and the data in the description file is converted into a mining format through the decomposition and merging of the data, the noise data is removed, and it is saved in the data warehouse. The inconsistency of personalized data is collected, the data that has nothing to do with mining requirements is deleted, and only collected data with html and htm suffixes is retained. The heuristic rules are used to identify users’ different IP addresses and divide access records into individual session fragments. In chronological order, the missing path of personalized data is supplemented. When users access English teaching resources, the personalized data is completed to form a complete conversation process [4]. Finally, a relatively complete conversation is formed, and then through the time window method, the logical unit for users to search for English teaching resources is obtained, and all the browsed teaching resources are merged into one transaction, and the small-grained transaction of personalized data is defined. When the preprocessed data set contains a large number of rule patterns, statistics and data mining methods are used to conduct in-depth analysis of the rule patterns, extract teaching resource rules and patterns that can represent user characteristics, and realize user search resource request identification. [5]. So far, the mining of user personalized data is completed.

2.2. Extracting keywords for teaching resources based on artificial intelligence technology
In the user's personalized data, the keywords of the text file are extracted and used as a keyword collection of English teaching resources. First, using artificial intelligence technology, according to the logic unit of the user's search for English teaching resources, the intelligent behavior of the user's search for resources is simulated, and the user's search resource terms are assumed to obtain key sentences and keywords with high frequency. Then the keywords and key sentences are decomposed to obtain single words with higher frequency. POSTag is used to improve the accuracy of single word judgment and the stop words in keywords are removed. Then the documents of English teaching resources are marked to construct an expanded map of teaching resource keywords, and the keywords and key sentences are used as a point in the map [6]. The details are shown in the figure below:

![Diagram](image)

**Fig.1** Teaching resource keyword expansion map

As shown in the figure above, the circles in the figure are key sentences with high frequency, and the rectangles are key words. Key sentences and keywords are connected to each other. Each key sentence has a limited number of key words, and some words are not included in the sentence. The similarity between keywords is used to obtain the semantic similarity of different keywords and encode the syntactic and semantic information of English teaching resources. According to the semantic factors of keywords, the random value of the content extracted by teaching resources is reduced, the trained public word vectors is used, the keywords are treated as node information to measure the semantic relationship between two nodes in the space. Describing the relationship between words and words, constructing an undirected graph, and using keywords as a node in the undirected graph, when keywords are contained in key sentences, the weights of different words are calculated to represent the similarity of two nodes, so that there is an undirected edge between different words [7]. Then the weight calculation formula $w$ between two word nodes is:

$$w = \frac{r}{m} \log \frac{t}{g+1}$$  \hspace{1cm} (1)

Where, $r$ is the number of individual words in the key sentence, $m$ is the number of keywords in the key sentence, $t$ is the number of sentences with a higher frequency, and $g$ is the word frequency of the collected reverse document. The weight value of each node is saved and normalized, and keywords are sorted and extracted according to the size of the weight value. Selecting keywords of English teaching resources, connecting them with other words, and judging sentences that can represent teaching resources, using two column vectors to iterate on the weights of keywords to obtain the relative contribution of the specified key sentences and keywords of the teaching resources, selecting the words with higher relative contributions, and using them as the keywords of the English teaching resources [8], at this point, the extraction of keywords for teaching resources is completed.
2.3. Recommending English teaching resources

The teaching resources that contain the keywords are recommended by generating user characteristics based on personalized data, calculating the similarity between keywords and user characteristics, and selecting keywords that meet user characteristics. User characteristics are divided into static attribute characteristics and dynamic attribute characteristics. Dynamic characteristics describe the user's dynamic behavior and obtain the data generated in the dynamic behavior. Static characteristics select the user's preference information and personal information. According to the relative contribution of teaching resource keywords, the keywords that can express user interests are selected, the characteristic values of users and teaching resources are obtained, the English teaching resource tags corresponding to the keywords is also obtained, as well as the initial English teaching resource recommendation list is obtained through offline calculation. The content recommendation methods are used to describe the attributes of English teaching resources, matching teaching resources with keywords representing user characteristics, and calculating the similarity between keywords and user characteristics. The similarity calculation formula is:

\[ Q = \frac{wL}{h} \]  \hspace{1cm} (2)

Where, \( Q \) is the similarity between keywords and user characteristics, \( L \) is the attribute value of the teaching resource, and \( h \) is the distance from the attribute value to the set of user needs. The greater the similarity, the more matching the teaching resource containing the keyword with the user, which can meet the individual needs of the user [9]. Calculating the similarity between all teaching resources and the user's characteristic keywords, the teaching resources with high similarity are taken as the recommendation target to be recommended to the user.

Interacting with users, showing users the recommended results of English teaching resources through the recommended list, and storing the user’s interactive behavior in the log, in this process, the appropriate recommendation algorithm is selected to further complete the collection and storage of user data, different recommendation engines are also selected. According to the various characteristics of users, the recommendation modules with different characteristics are divided. Each recommendation engine is responsible for the recommendation of a user characteristic, and the LOM standard description of the English teaching resources that need to be recommended is described, so that each engine represents a recommendation strategy. The emphasis of user characteristics are predicted, the weight values of different recommendation engines are determined, the judgment of the user's personalized characteristic tendency is realized, and the priority level of the recommendation engine is determined. Oriented by the user’s learning interest and style, through the way of resource presentation, the English teaching resources described in the LOM standard are converted, the format of the teaching resources required by the user is judged, and the text title keywords, video icons and audio of the teaching resources are distinguished. Through Web crawler technology, English teaching resources related to recommended resources are obtained, and the recommended formats of keywords, video and audio are selected [10]. Mapping recommended resources to implicit classifications, automatically clustering user characteristics, labeling English teaching resources and users, and realizing the connection between English teaching resources and users through hidden characteristics, and completing the implicit semantic classification of teaching resources, by selecting a good recommendation format, all the English teaching resources in this category are jointly recommended to users. At this point, the design of a personalized recommendation method for English teaching resources based on artificial intelligence technology is completed.

3. Experimental demonstration analysis

A comparative experiment was carried out. The recommended method was recorded as experimental group A, and the traditional method was recorded as experimental group B. The accuracy of the two
methods recommended for English teaching resources was compared, and the average absolute deviation is used as the accuracy evaluation standard. The experimental configuration is Intel Core 2 dual-core processor, the online recommended platform for teaching resources is MyEclipse 6.5, and the programming language is Java. The English teaching resources collected by the platform are shown in the following table:

| Specialty code | Resource category          | Number of resources |
|----------------|---------------------------|---------------------|
| 100            | English words             | 95                  |
| 130            | Analysis of English Textbooks | 47                 |
| 140            | English grammar           | 95                  |
| 160            | Oral and Translation      | 77                  |
| 210            | Aural comprehension reading | 99                 |
| 300            | English reading           | 120                 |
| 510            | English Magazine          | 48                  |
| 620            | English essay             | 48                  |
| 710            | Business English          | 99                  |
| 820            | IELTS English             | 46                  |

Tab.2 Classification of teaching resources for datasets

Two sets of experiments are conducted on 5 users, the data set is divided into training set and test set, and the mean absolute deviation MAE is selected as the evaluation standard of recommendation accuracy. The definition formula $MAE$ of the mean absolute deviation is:

$$MAE = \frac{1}{n} \sum_{i=1}^{n} |p_i - q_i|$$

Where, $p_i$ recommends a set of ratings for teaching resources, $n$ is the number of ratings, $i$ is the number of sets, $q_i$ is the corresponding set of actual user ratings, the smaller the MAE value, the higher the recommendation quality of English teaching resources. The collected data set is used to adjust the nearest neighbor size of the training set and the test set to make the two sets of experiments produce the best results in the recommendation process. The accuracy of the recommended results of the two sets of experiments is compared and the MAE values under different nearest neighbor numbers are recorded. The comparison results are shown in the following table:

| Number of nearest neighbors | MAE value of experimental group A | MAE value of experimental group B |
|-----------------------------|----------------------------------|----------------------------------|
| 3                           | 0.79                             | 0.87                             |
4 0.81 0.76
5 0.80 0.83
6 0.82 0.82
7 0.81 0.84
8 0.83 0.84
9 0.82 0.87
10 0.80 0.88
11 0.81 0.85
12 0.79 0.88
13 0.82 0.87
14 0.82 0.85
15 0.80 0.86

Tab.3 Experimental comparison results

As shown in the above table, when the number of nearest neighbors is different, the MAE value of experimental group A is lower than the MAE value of experimental group B, the average MAE value of group A is 0.81, and the average MAE value of group B is 0.85. The MAE value of group A decreased by 0.03 on average. In summary, this method adopts artificial intelligence technology and makes full use of personalized information data to simulate and predict the similarity of keywords. Compared with the traditional recommendation method, it can ensure the accuracy of teaching resource recommendation, effectively improve the quality of resource recommendation, make the recommendation result more in line with the learning needs of users, and improve user experience.

4. Conclusion
The method proposed in this paper takes full advantage of artificial intelligence technology, reduces the MAE value of the resource recommendation algorithm, and improves the accuracy of resource recommendation. However, there are still shortcomings in this research, such as insufficient sharing of English teaching resources. In future research, collaborative filtering and data mining technology will be combined in the recommendation process, and the open Internet will be used to introduce more off-site English teaching resources.

References
[1] Susheng Z, Ming L and Lei Yue Y 2019 Research on the Recommendation Strategy of Private University Teaching Service Based on Multi Feature Fusion Modern Information Technology 3(24) 118-120
[2] Ting H 2019 Design of English Teaching Resource Management System Based on Collaborative Recommendation Techniques of Automation and Applications 38(09) 158-161
[3] Yi H 2019 The Construction of Online Educational and Teaching Resources Recommendation Model Based on Theme Words I Techniques of Automation and Applications 38(09) 170-173
[4] Yanhong Z and Hang S 2019 Research on Label-based Recommendation Algorithms for Self-help E Education Resources Modern Information Technology 3(12) 13-17
[5] Shuangxiang Z Teaching Information Resource Management Method Based on User Interest in HDFS Mode Modern Electronics Technique 42(11) 87-89
[6] Wenxin L, Yongjun W and Lijun T 2019 Research and Implementation of Individualized Recommendation Methods of Educational Resources COMPUTER TECHNOLOGY AND DEVELOPMENT 9(06) 18-22
[7] Tingting L and Liqin L 2018 Analysis of the Personalized Recommendation Algorithm and Platform Design of Learning Resources Education Modernization 5(52) 171-174
[8] Huawei Z Personalized Teaching Resource Recommendation System Based on LDA Modern
[9] Yufei L and Shen J 2020 Design of Personalized and Precise Recommendation System for College Students' Ideological and Political Education Resources *MICRO COMPUTER APPLICATIONS* **36**(08) 55-58

[10] Wei W 2020 Design of an Intelligent Recommendation and Retrieval System for Digital Archives Resources Based on Attribute Grouping *Electronic Design Engineering* **28**(15) 19-28