

An Algorithm for Generating a Recommended Rule Set Based on Learner's Browse Interest

<https://doi.org/10.3991/ijet.v13i11.9604>

Xiaowei Hao^(✉), Shanshan Han
Hebei University of Science and Technology, Hebei, China
haoxiaowei12930@126.com

Abstract—To personalize the recommended learning information according to the interests of the learner, a recommendation rule set generation algorithm based on learner browsing interests was proposed. First, the learner's browsing behavior was captured. A multivariate regression method was used to calculate the quantitative relationship between the learner's browsing behavior and the degree of interest in the web page to generate a learner's current interest view (CIV). With this current interest view, a content-based collaborative filtering personalized information recommendation service was provided to learners. Then, a new weighted association rule algorithm was used to discover the associations between the items, so that the degree of recommendation was obtained. Furthermore, the degree of recommendation was used as a personalized recommendation service for learners with long-term interests. The results showed that the proposed algorithm effectively improved the quality of information recommendation and the real-time performance of the recommendation. Therefore, this algorithm has a good application value in the field of personalized learning recommendation.

Keywords—weighted association rule algorithm, browsing interest, personalized recommendation, current interest view

1 Introduction

With the development and popularization of the information superhighway, the learning style is constantly changing. The information resources on the Internet are exponentially expanding and are the source of massive information. The organization of its information is heterogeneous, diverse, and distributed. The Web is an important way for learners to access information. However, due to the growing information, learners have to spend a lot of time looking for information of interest. Even if some valuable information is found, there is often a lot of "noise" mixed. The search engine is a universal tool that assists learners in retrieving information. Traditional search engines such as AltaVista, Yahoo and a new generation of search engine Google and other information retrieval technologies meet the needs of learners. Due to its general nature, query requests of different backgrounds, different purposes and different periods cannot be met. Therefore, how to enable learners to quickly and effectively obtain

information of interest, and how to actively recommend the information of interest to learners to learn the user's service has become a common concern of many users and learning platform operators.

Personalized services are addressed to these issues. It provides different services for different users to meet different needs. Personalized services learn the interests and behaviors of users by collecting and analyzing user information, thereby achieving the purpose of actively recommending information. It transforms the Internet from passively accepting viewer requests to actively perceiving viewers' information needs. The active messaging service for the viewer by the Internet system is implemented. The Internet personalized recommendation system is a hot issue in the research of e-commerce, Internet technology and data mining at home and abroad. At the same time, the existing recommendation techniques have some shortcomings. For example, rule-based recommendation techniques cannot be dynamically updated. As the number of rules increases, the system becomes more and more difficult to manage. The recommendation technology based on information filtering also encounters two difficult problems in the practice process, namely sparsity and scalability. Moreover, as the system continues to grow, the existing recommendation algorithms also expose many shortcomings. The recall rate and accuracy rate of the recommended results are still not satisfactory. Therefore, the research of personalized recommendation technology has academic value.

2 State of the art

The method of providing information recommendation to users is studied, which can help users find valuable information. It has gradually become an important part of personalized service research and has been paid more and more attention by researchers. Hsu et al. [1] proposed a mobile language learning method based on personalized recommendation. Based on this method, a mobile learning system was developed. By providing a reading material recommendation mechanism, it directs EFL (English as a Foreign Language) students to read articles that match their preferences and knowledge levels. An annotation module that enables students to read is designed. Salehi and Kamalabadi [2] proposed a new material recommendation system framework based on sequential pattern mining and multi-dimensional attribute collaborative filtering. In the sequential pattern-based approach, the improved Apriori algorithm and the PrimeSpriSn algorithm are implemented to discover potential patterns in material access and use them for recommendation. The research shows that this method is superior to the previous classification accuracy measurement algorithm and can accurately satisfy the learner's real learning preference according to the context information updated in real time. Wang et al. [3] proposed a new method of music recommendation. By balancing this exploration and development, learning tasks are strengthened. To study user preferences, a Bayesian model is used. Suggestions for audio content and novelty were presented. Piecewise linear approximation models and variational inference algorithms help accelerate Bayesian inference. Dorça et al. [4] proposed an effective method for individualizing the teaching process based on learn-

ing style. Based on the expert system, this method implements a set of rules for classifying learning objects according to the teaching style. Then, based on the student's learning style, the learning object is automatically filtered. The best adapted learning subject rankings and recommendations are provided to students. Wu and Chen [5] propose personalized recommendation for E-learning system based on social network analysis. Shivanagowda et al. [6] analyzed the data generated by students' activities in the engineering and technical resource compiler. These data are indeed key useful data for the teacher/teaching system in generating personalized recommendations, which can change the students who are limited to video resource learning. Yang [7] believes that information overload is the main factor affecting the performance improvement of mobile learning. From the perspective of personalized service, the mobile learning resource recommendation model is constructed for the individualized needs of mobile learning resources, contextualization, and intelligence, and the key technologies of personalized resources are analyzed. Guo and Zhao [8] proposed a combined recommendation model. Content filtering recommendation technology and collaborative filtering recommendation technology based on mainstream personalized recommendation technology are analyzed and compared, and the working process of the model is given.

In summary, most scholars have studied the establishment of personalized recommendations, and most of them are implemented through filtering recommendation techniques. Using the rule set generation algorithm, the learner's personalized learning is realized, which is innovative. It also has reference value for the future development of personalized recommendations.

3 Methodology

3.1 Personalized recommendation model based on weighted association rules and browsing behavior

According to the shortcomings of the traditional personalized recommendation process, a new personalized recommendation model based on weighted association rules and browsing behavior is proposed. The model is shown in Figure 1:

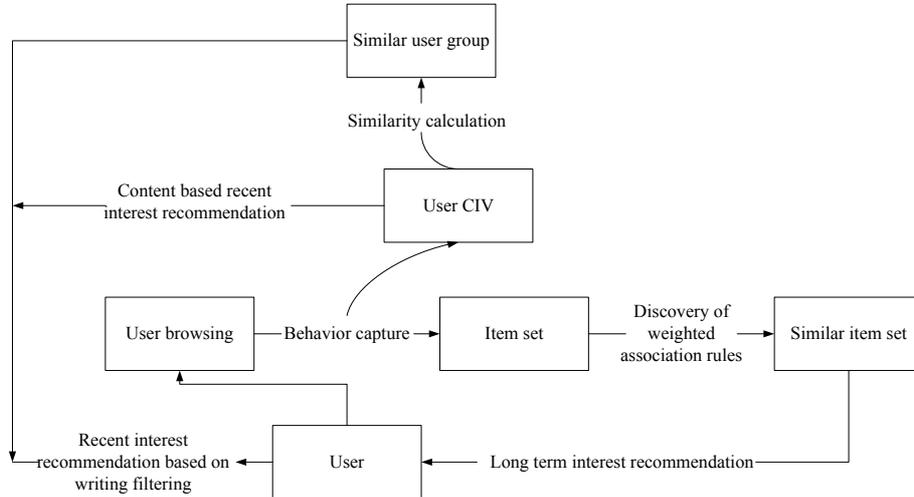


Fig. 1. New personalized recommendation model

3.2 Current interest view based on learner browsing behavior

Obviously, the learner's interest is closely related to the browsing behavior of the web page being viewed. On the surface, it is possible to reveal that the learner's interest in the web page P is a lot of browsing behavior $d(P)$. However, the following two actions play a key role. One is the browsing time $t(P)$ on the web page P (referred to as the T behavior), and the other is the page turning/pulling scroll bar number $v(P)$ (referred to as the V behavior).

In order to find the quantitative relationship between T and V behavior and web interest, careful analysis and experimentation were carried out. Multiple linear regression methods are used as tools for web page interest modeling analysis. $d(P)$ is a random variable related to $t(P)$ and $v(P)$. For each set of values of $t(P)$ and $v(P)$, there are:

$$d(P) = a * t(P) + b * v(P) + c + \varepsilon \tag{1}$$

In the formula, a , b , σ^2 are unknown parameters that are independent of $t(P)$ and $v(P)$. ε is a random error, which obeys the normal distribution $N(0, \sigma^2)$. It is a multivariate normal linear regression model.

$$d(P) = a * t(P) + b * v(P) + c \tag{2}$$

a and b are called regression coefficients (a , b , and c are behavioral influence factors). The parameters a, b, σ^2 can be estimated by least squares.

The recent interest of learners refers to the interest of users in the Internet for the most recent period of time. CIV (Current Interest View) is used to describe. The CIV

generation process is divided into two phases. The first stage is to hierarchically cluster the web pages that the learners have recently browsed to obtain the Web Pages Classification Tree (WPCT). WPCT can roughly describe the learner's browsing interests. In the second stage, learner interest is not only related to the content of the webpage being viewed, but also to the learner's interest in each webpage. The degree of interest of each web page in WPCT, the interest level of each category, and the interest density of each category are further calculated to obtain CIV.

The learner interest discovery and the interest description model were selected according to the standard category tree (SCT). Web page classification tree (WPCT) shows learners' interest in subclasses and their degree of interest in standard category tree. After determining the standard category tree, the mining target is positioned to "discover the learner's sub-categories of interest in the standard category tree and their level of interest" to form WPCT with standard classification as nodes. It can roughly reflect the classification and extent of interest to learners. The mining will be based on the text content of the webpage that the learner has viewed. The mining strategy will adopt a top-down hierarchical taxonomy (according to SCT). The sequence algorithm is chosen in the algorithm design, which is not only easy to implement, but also improves the accuracy of classification by means of semantic analysis. In WPCT, although $n(T_i)$ can be directly used to describe the learner's interest in classifying T_i , this description is not accurate enough. If the learner has viewed 3 pages in the data mining category, the interest level of each page may be only 0.4, and his interest in 2 pages in the artificial intelligence classification may be 0.8 and 1, respectively. Therefore, the interest of learners in these two categories should be 1.2 and 1.8 respectively. Although the $n(T_i)$ value of the latter classification is small, learners are more interested in this classification. To overcome this phenomenon, the impact of the number of web pages in the classification and the interest of each web page on the classification interest will be considered. The concepts of classification interest degrees $d(T_i)$ and $id(T_i)$ are introduced.

$$d(T_i) = \sum_j d(P_{ij}) \quad P_{ij} \in T_i \quad (3)$$

$$id(T_i) = d(T_i) / Leaves(T_i) \quad (4)$$

In the formula, $Leaves(T_i)$ represents the number of leaf nodes of the subtree obtained from the classification T_i in the SCT.

CIV is a tree structure whose structure is very similar to WPCT, but all web nodes are removed and only the classification nodes are retained. Each node is a triple $(T_i, d(T_i), id(T_i))$, which consists of a classification name, a classification interest, and a classification interest density. The generation of CIV is done by the client's CIVG agent, and the generation process is divided into two phases. The first is to hierarchically classify the web pages that have been viewed recently to obtain WPCT. The second is to calculate the interest degree of each web page in WPCT, the interest degree of each category and the interest density of each category, to obtain CIV. The steps of the CIV generation algorithm are as follows: First, WPCT is generated ac-

ording to the web page set browsed by the user. Then, each of the WPCTs is classified as T_i . According to formula (3) and formula (4), the interest degree, $d(T_i)$ and $id(T_i)$ of each web page under T_i are sequentially calculated. Finally, the user CIV can be obtained based on the WPCT structure and the calculated classification interest and interest density.

3.3 Nearest neighbor concern project discovery based on weighted association rules

To reflect the importance of different projects in the item set and the association between the mining projects, the existing problem model is extended, and a new so-called weighted association rule problem is proposed. To find the weighted association rules in the database, the k -support expectation concept of the project is proposed. Based on this, a discovery algorithm for weighted association rules is proposed.

The input data is expressed as a $R:m \times n$ user-item selection matrix. If the user selects an item, the user is concerned about the item. m is the number of users, and n is the number of items. R is a matrix, and $r_{ij}=1$ means that the i -th user has selected the j -th item. $r_{ij}=0$ means that the item is not selected. The collection of items is $I=\{i_1, i_2, \dots, i_n\}$. To characterize the importance of different items in the item set, each item ij is assigned a weight w_j . When assigning weights to items, in order to reflect the degree of attention of the project, the frequency selected by the user is used as the weight of the item. The number of users of the item ij is N_j , $w_j=N_j/m$, and it is taken as an integer. In addition, the concept of a user profile is introduced. The profile of user i is represented by a row vector U_i . The elements in U_i are the items that the user has selected. The set of vectors U_i is the set of all user profiles, which is denoted by U . The algorithm for the item weighting association rule in the user profile set is as follows:

Input: user profile set, weight W_j for each item, minimum confidence min_conf , minimum weighting support ω_min_sup ; frequent item set $L=\varnothing$;

Output: weighted association rule set WAR; // The value of the largest length of the user profile set vector U_i is taken as the maximum possible length of the frequent item set, i.e., k - supports the maximum value of the expected k .

```

Sizemax=Max(size( $U_i$ )); //  $1 \leq i \leq m$ 
for  $i=1$  to Sizemax do
     $C_i=\varnothing, L_i=\varnothing$ ; //  $C_i$  is a collection of possible frequent  $i$ -project subsets,  $L_i$  is a collection of frequent  $i$ -items. Their initial collections are all empty.
    for  $j=1$  to  $n$  do
        { $sc(i_j)$ ; // the process  $sc(i_j)$  calculates the number of supports for  $ij$ , that is, the number of users who have selected  $ij$ .
        for  $k=2$  to Sizemax do
            if  $sc(i_j) \geq B(\{i_j\}, k)$  then
                 $C1 \leftarrow i_j$ ; // a set of possible frequent  $k$ -project subsets that form frequent 1-item sets
            }
    }

```


the density of interest in CIV are analyzed. By comparing the content, web pages that are closely related to learners on the Web site are screened out and recommended to learners. First, the individual learner CIV uses the breadth-first search method to select all the classes of interest of $id(T_i) > \alpha_{n_Interest}$ (that is, the threshold at which the learner is interested in the classification). It is represented by a vector IT, that is, $IT = \{T_1, T_2, \dots, T_k\}$. Each class of interest T_j of the learner is then determined to be a set of keywords representing the classification, i.e. a local dictionary of the classification, which may be derived from domain knowledge or a professional dictionary. The sequence algorithm is specifically designed to classify the text using a local dictionary, and the algorithm has a better effect on text classification. The web page can be viewed as a text file, and the learner interest classification is described in a local dictionary. According to T_j 's local dictionary, a sequence algorithm is used to filter out web pages related to T_j on the website. The similarity obtained by the sequence algorithm is sorted from high to low. Finally, for the learner's information recommendation, the top-N recommendation is used to recommend the top N items with the highest similarity to the learner. In general, learners' attention is mainly concentrated in the first seven items. Therefore, the N of the top-N recommendation algorithm generally has a value of 10.

4 Result analysis and discussion

4.1 An experiment on the long-term concern of the learner

The 400 pages of the computer knowledge topic on the <http://it.163.com> website were downloaded and saved on the experimental PC. For ease of operation, each web page is provided with an index number. When programming, the index number represents each web page, which facilitates the implementation of three algorithms, and the efficiency of association rule mining and nearest neighbor discovery is improved. On the experimental PC, three methods of learner's attention item recommendation algorithm (WIR) based on weighted association rules, traditional association rule recommendation algorithm (ARR) and project evaluation based collaborative filtering recommendation algorithm (IRP) are implemented in Java language. From the history information of the 400 web pages accessed by the learner, the access information of 20 learners is selected. Each of the 20 learners visited at least 20 pages.

In the experiment, the average click rate of the recommended items is used by the learner to measure the accuracy of the recommendation algorithm. The recommended project should meet the needs of the learner. If the accuracy of the recommendation is high, the learner's click rate is high. In the test data set, the total click rate of each learner is divided by the number of learners to get the average click through rate. The hit rate of learners is equal to the number of actual clicks of learners divided by the number of items recommended. Studies have shown that learners are generally more interested in the first seven items recommended for it, and that learners after 7 will not see it, so the maximum number of items recommended in the experiment is 10. When the recommended number of pages is $C=2, 4, 6, 8, 10$, the learner's average click rate is calculated separately, as shown in Table 1 and Figure 2.

Table 1. Average click rate

Recommended number of pages C	WIR	ARR	IRP
2	0.3755	0.3250	0.3198
4	0.4124	0.3874	0.3774
6	0.4244	0.4083	0.3982
8	0.5182	0.4994	0.4874
10	0.5556	0.5257	0.5238

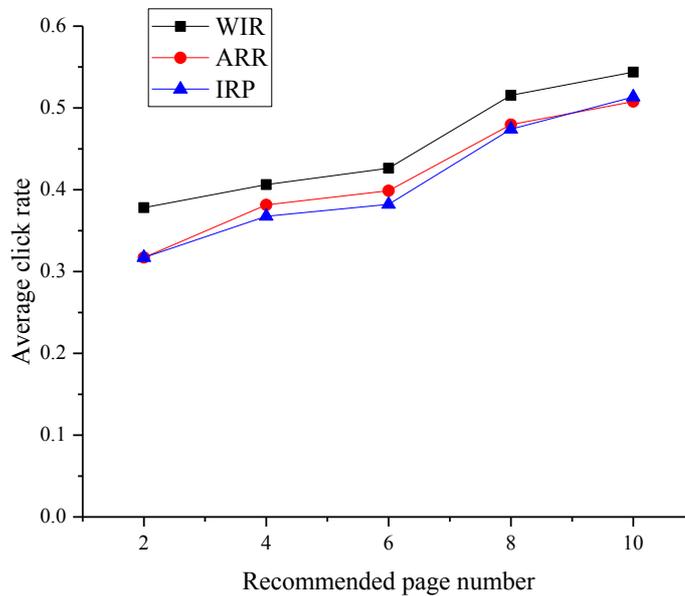


Fig. 2. Average click rate

From Table 1 and Figure 2, it can be seen that the user's click rate (that is, the accuracy rate) of the proposed WIR algorithm is higher than that of the ARR and IRP algorithms. The user-click rates of the ARR and IRP algorithms are similar.

When comparing the performance of an algorithm, the average prediction time is used to measure. The average prediction time is represented by the sum of the training time and the test time (corresponding to the learning phase and the application phase of the recommendation system, respectively) divided by the number of predicted items. The performance comparison of the three algorithms is shown in Table 2 and Figure 3 below. In Figure 3, the abscissa is the number of item set items (web pages), and the ordinate is the average predicted time. The average forecast time is in seconds.

Table 2. Average forecast time

Number of pages	WIR	ARR	IRP
20	90.467	90.297	90.802
40	100.212	98.792	100.641
60	120.368	116.842	130.478
80	125.494	120.274	150.721
100	140.197	129.858	180.131

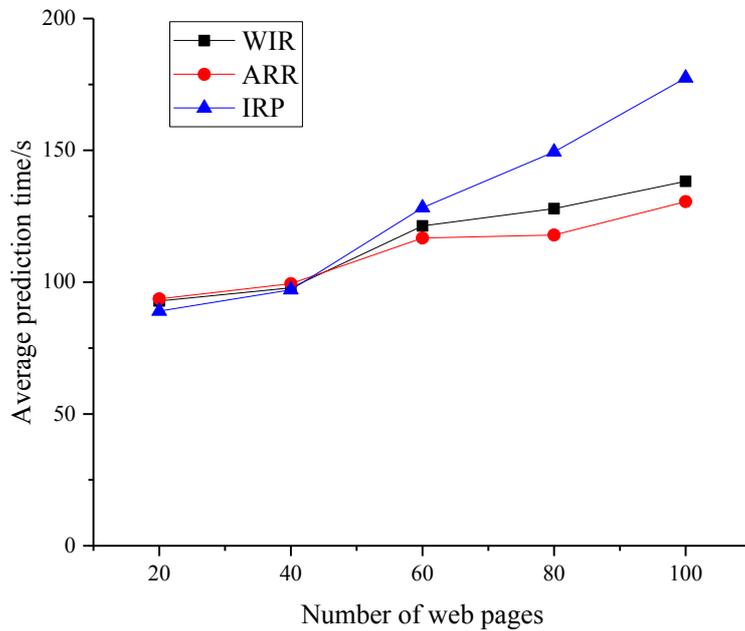


Fig. 3. Average forecast time

It can be seen from Table 2 and Figure 3 that when the number of web pages is small, the average prediction time of the three algorithms is similar. However, when the number of web pages increases gradually, the average prediction time using the IRP algorithm becomes more and more obvious than the average prediction time using the other two algorithms. The reason why the proposed average prediction time of the WIR algorithm is slightly larger than that of the ARR algorithm is that the WIR algorithm introduces the selection attention in the recommendation. The product of the degree of interest and the degree of confidence is used as the degree of recommendation. Then, according to the size of the recommendation, the item is recommended. The ARR algorithm directly recommends the confidence as the size of the

recommendation. Therefore, it takes a certain amount of time to calculate the recommendation degree in the WIR algorithm.

Through the comparison of the accuracy and performance of the above three algorithms, it is found that the WIR algorithm has the highest accuracy. When recommending for learners, the first thing to consider is the quality of the recommendation. The performance of the WIR algorithm is only slightly lower than the ARR algorithm. Therefore, the proposed WIR algorithm is a recommendation algorithm with good quality and efficiency.

4.2 Recommendation based on the learner's current interest view

In the recommendation experiment based on recent learners' interest, a learner's browsing page and browsing behavior data on the site are used to illustrate the method of current interest view of the learner, which is obtained by using the regress command in the Matlab to solve and test the regression equation, the calculated web page interest and the web page classification tree. Then, the behavior data of the other three learners was collected. Using the current interest view of the three learners, the project is recommended. The recommended average recall rate and accuracy are calculated.

First, for the computer knowledge topic under this website, the local dictionary of all the keywords needed for the experiment is extracted. Text files are used to record data. On this basis, the standard category tree (SCT) is formed for the computer knowledge topic under this website, as shown in Figure 4.

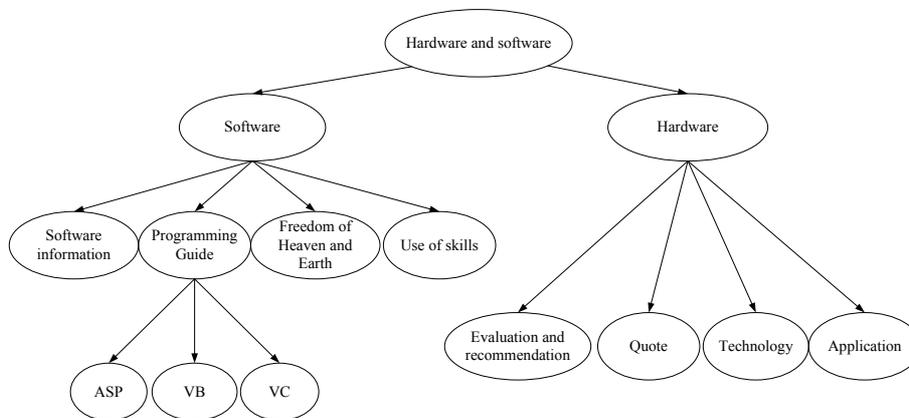


Fig. 4. Standard category tree

Through the browser, the behavior data of a learner arbitrarily browsing 200 web pages is captured. Using the regression equation and calling the regress command provided by Matlab, the empirical regression equation can be obtained:

$$y = 0.1300x_1 + 0.1017x_2 + 62.0492 \quad (6)$$

The test regression equation is used to calculate the interest level of the web page, and its rationality is checked. In the experiment, 20 web pages in the web pages that the learners browse randomly are evaluated. The estimated value is compared with the calculated value. The results show that the values are very close. It is reasonable to use the regression equation to calculate the interest level of the webpage.

Then, the method of classifying the interest degree and classifying the interest density value is used to calculate the value of the number of classified web pages, the classification interest degree, and the classification density when the learner browses 200 web pages, as shown in Figure 5. The learner's classification interest vector density can be obtained, and the learner is most interested in the classification VB.

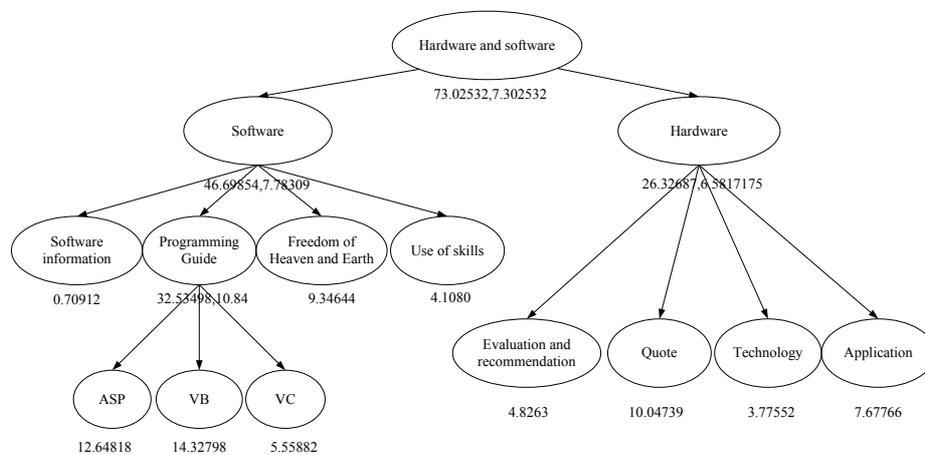


Fig. 5. Learner's current interest view

Using the same method, the CIV of the learner browsing 50, 100, 150, and 250 web pages can be calculated. At the same time, in order to verify the universality of the method, in addition to the behavior data of the above learners, among the learners participating in the experiment, two learners were randomly selected for testing. Their behavior data was recorded when they browsed 50, 100, 150, 200, and 250 pages, respectively. In the same way, the CIV of the two learners arbitrarily browsing 50, 100, 150, and 250 pages was generated. Using the five sets of experimental data of each of the three learners, the precision, recall, and average of the recommendations were calculated. Precision = Recommended number of correct messages / actual recommended number of messages. Recall = Recommend the correct number of messages / the number of messages that should be. The average recall rate recommended for the three learners is shown in Figure 6 and the accuracy is shown in Figure 7. In the two figures, the abscissa indicates the number of web pages viewed by the learner when calculating the learner's behavior parameters. The ordinate in Figure 6 indicates the recall rate at the time of recommendation. The ordinate in Figure 7 indicates the accuracy used for the recommendation.

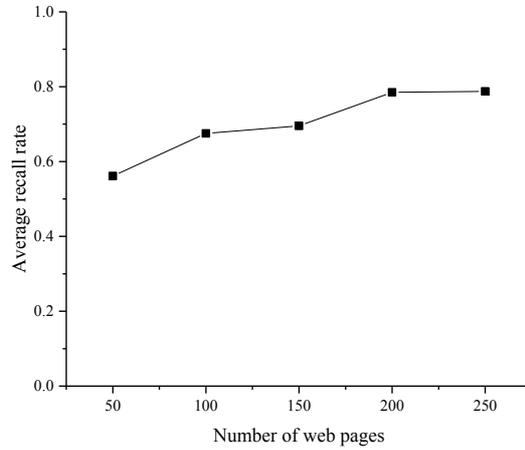


Fig. 6. Average recall rate

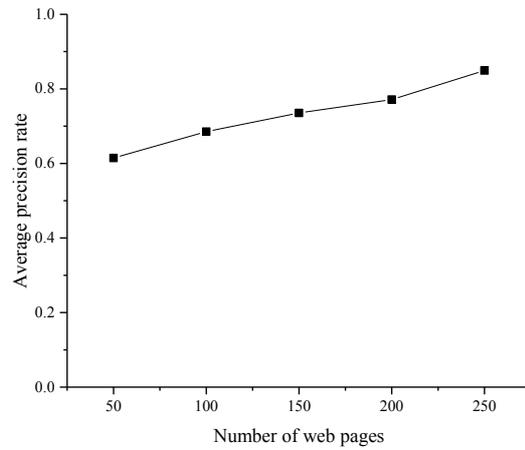


Fig. 7. Average accuracy

It can be seen from Figure 6 and Figure 7 that the average value of the recommended recall rate using the CIV model is 72.8%, and the average accuracy rate is 69.8%. Considering some human factors that cannot be overcome, recall rate and accuracy rate are still ideal. From the above two figures, it can also be seen that the more the number of web pages used to calculate the behavior parameters, the higher the recall rate and accuracy rate. Therefore, it can be concluded that when there are many web pages viewed, the learner's behavior parameters calculated by multivariate linear regression are accurate, which is in accordance with the statistical principle.

5 Conclusions

When recommending information according to the long-term interest of learners, a new learner attention item recommendation algorithm based on weighted association rules is proposed. When recommending recent interest information for learners, a new personalized recommendation method based on learner current interest view is proposed. Finally, the accuracy and performance of the recommendation algorithm, the traditional association recommendation algorithm based on association rules, and the filtering recommendation algorithm based on the collaboration of nearest neighbor learners are compared. The experiment gives the accuracy and recall rate when using the learner current interest view for personalized recommendations. The validity of the algorithm was tested. The proposed algorithm has a good application value in the field of personalized learning recommendation.

6 References

- [1] Hsu, CK., Hwang, GJ., Chang, CK. (2013). A personalized recommendation-based mobile learning approach to improving the reading performance of EFL students: *Computers & Education*, 63(1): 327-336 <https://doi.org/10.1016/j.compedu.2012.12.004>
- [2] Salehi, M., Kamalabadi, IN. (2014). Personalized recommendation of learning material using sequential pattern mining and attribute based collaborative filtering: *Education & Information Technologies*, 19(4): 713-735 <https://doi.org/10.1007/s10639-012-9245-5>
- [3] Wang, X., Wang, Y., Hsu, D., Wang, Y. (2014). Exploration in Interactive Personalized Music Recommendation: A Reinforcement Learning Approach: *Acm Transactions on Multimedia Computing Communications & Applications*, 11(1): 1-22
- [4] Dorça, FA., Araújo, RD., De Carvalho, VC., Resende, DT., Cattelan, RG. (2016). An Automatic and Dynamic Approach for Personalized Recommendation of Learning Objects Considering Students Learning Styles: An Experimental Analysis: *Informat-ics in Education*, 15(1): 45-62
- [5] Wu, B., Chen, PP. (2013). Personalized Recommendation Research in E-Learning Systems: *Applied Mechanics & Materials*, 433-435: 603-606 <https://doi.org/10.4028/www.scientific.net/AMM.433-435.603>
- [6] Shivanagowda, GM., Goudar, RH., Kulkarni, UP. (2015). nalysis of New Data Sources in Modern Teaching and Learning Processes in the Perspective of Personalized Recommendation: *Smart Innovation Systems & Technologies*, 31: 529-539 https://doi.org/10.1007/978-81-322-2205-7_49
- [7] Yang, L., Yang, Y. (2015). Research for Personalized Recommendation of Learning Resource in Mobile Computing Context: *Lecture Notes in Electrical Engineering*, 352: 309-315 https://doi.org/10.1007/978-3-662-47487-7_44
- [8] Guo, JX., Zhao, JC. (2013). Research on Personalized Courseware Recommendation System of Rural Distance Learning Based on Combination Recommendation Technology: *Applied Mechanics & Materials*, 373-375: 1652-1660 <https://doi.org/10.4028/www.scientific.net/AMM.373-375.1652>

7 Authors

Xiaowei Hao and **Shanshan Han** are with Hebei University of Science and Technology, Hebei 050018, China (haoxiaowei12930@126.com).

Article submitted 27 September 2018. Resubmitted 28 October 2018. Final acceptance 13 November 2018. Final version published as submitted by the authors.