

# Micro Course Recommendation Model Optimization based on User Behavior

Xiangbin Gao<sup>1, a</sup>, Lian Wang<sup>2</sup>

<sup>1</sup>Linyi University, Shandong 273400, China

<sup>2</sup>Shandong Feixian Experimental Primary School, China

<sup>a</sup>gxb0715@126.com

**Keywords:** Recommendation Model; Search Engine; Micro Course; User Interest.

**Abstract.** With the advent of the information era, it has brought a change to the education industry. As the network educational resources gradually increased, to obtain individual needs in the massive resources has become a headache problem for users. In the use of traditional search engines for information retrieval, users need to actively input the retrieval keywords, and the search results have no users individualization feature. In this case, recommendation technology has emerged as a complementary technique for search engine. In this paper, we combined recommendation technology with micro course system in the education field, through the establishment of user interest model, at last made use of the recommendation algorithm based on users as the bridge to link the user interest and micro course video resources, presented the video resources that users are satisfied with and interested in to the user, and completed personalized video recommendation. The combination of the recommendation model and the micro course has a positive effect on reducing the time needed to retrieve the video resources and improving the learning and working efficiency of teachers and students.

## 1. Introduction

In the micro course recommendation model, user interest model plays a very important role in the recommendation results quality, and all the recommendations take the user interest model as the foundation. Because the system can not only understand the needs of users, but also need to completely and accurately express this demand based on the understanding. If the preference got is consistent with the reality, but the problem appears in the expression process, then the negative effect will be reflected in the user satisfaction of the system [1-2]. Usually, we will combine with the actual situation, and make use of a reasonable data structure to represent the user interest model. We require this data structure can not only express the user interest model accurately, but also convenient to do some calculation according to the model. Currently, the commonly used method is some model expression methods in the field of information retrieval. These traditional expression methods are mature, able to accurately express and meet our expression needs [3].

In this paper, we combined recommendation technology with micro course system in the education field, through the establishment of user interest model, at last made use of the recommendation algorithm based on users as the bridge to link the user interest and micro course video resources, presented the video resources that users are satisfied with and interested in to the user, and completed personalized video recommendation.

## 2. User interest model expression methods

### 2.1 Vector space model expression method

Vector space model (VSM) was firstly used in SMART information retrieval system, which has become the most commonly used model in Natural Language Processing [4]. Compared to other model representation methods, this method is currently widely used and popular by a large number of users. Each vector in the vector space model consists of two parts information indexes: the interest item and the weight. First of all, make use of reasonable information feature item to express the needs

of the user's interest in the system. In addition, each keyword corresponds to the unique weight, and the weights represent the user interest level. The two parts of the information constitutes a vector. In the representation of user interest model, it is actually one or more of the vector. Each keyword in the model representation will be the basis for the recommendation of the system. But this method also has a more obvious drawbacks: each dimension of the vector is actually corresponding with the user's interest item. As a result, with the increase of user interest item, the model dimension will also continue to increase, then in the relevant calculation of data in the model, time spent and the space cost by the calculation will also increase the burden for the model calculation.

## **2.2 Keywords list representation**

Keywords list representation method is to get the key words that can express the user's interest, and use it to construct the user interest model [5]. For example, a video site, there is a user who is quite interested in the comedy movie, so the user interest about the user in the use of the method can be represented as {comedy, Ge You, Hollywood} this kind of form. But in terms of access to key words of user interest, an acquisition way is to actively fill in and give feedback to the system when users use the system. The other way is that the system obtains information through the historical analysis of user behavior data. At the same time, these key words will change with the dynamic time and other factors change. And the change does not occasionally happen, but always change continuously, and follows certain rules, cycling in a certain range of time. The most representative of the keywords list representation system is the Web Watcher recommendation system.

## **2.3 Score matrix representation between the user and items**

The score matrix representation between the user and the item is on the basis of  $m \times n$  dimension, in which  $m$  and  $n$  refer to user space and item space respectively. For each data item in the corresponding matrix, it is actually the score of a user for an item in the system. The evaluation numerical is proportional to the user interest degree of the item. Each row vector in the matrix corresponds to all projects evaluation of a user, therefore, the score can directly reflect the preferences of a user to the interest item. Usually, in the actual system applications, the system will provide different scoring levels for users to score or choose, and most of the current Internet scoring levels are ranged in  $[0,5]$  in the score.

## **2.4 Coarse and fine interest granularity representation**

The coarse and fine interest granularity representation method, when in use, it is necessary to combine the division of the user interest classification in the system, and according to the division scale, to eventually form different granularity in the model representation [6]. In the coarse-grained representation, the most common method is to divide into a simple like and dislike two interest criteria. The larger the divided particles, the relatively easier to achieve; in the fine-grained representation, the classification method is more specific. It will correspond the user interest to different granularity, and use various evaluation dimensions to represent the interest of the user. The all-round, comprehensive and detailed information representation and measure, although it will be difficult to achieve, the accuracy and intuitive get a great improvement. In practical applications, the coarse-grained representation is often used in a wider range because of the simplicity in realization.

## **3. User interest model establishment**

Preparation before the modeling, make planning for the following model establishment process [7-8]:

- (1) What are the input data of the model; where the input data are obtained; how to obtain them;
- (2) What are the objects to be modeled in the model;
- (3) How to model; what is the process of modeling;
- (4) After modeling, the output of the model and the representation of the model;
- (5) Whether the data in the model is static or not; how to learn and update the model.

In the following, based on the above model building ideas, we develop the corresponding research. The user modeling process diagram is shown in figure 1.

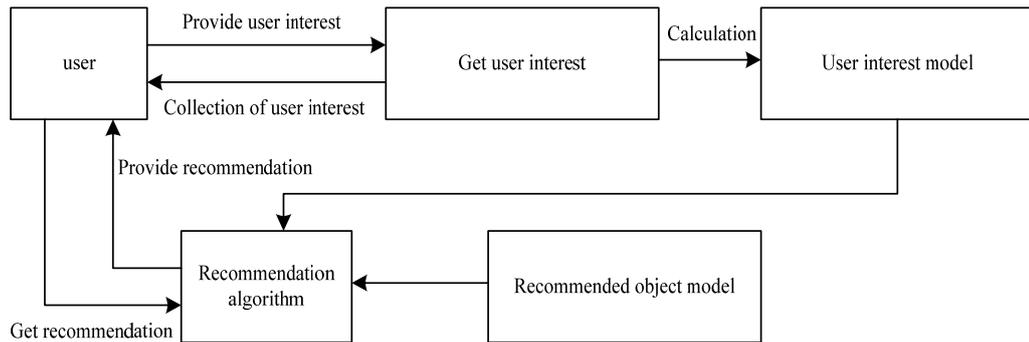


Figure 1. Common process of user modeling

### 3.1 Micro course feature analysis

(1) The user significant data acquisition. In the user explicit data, in this study, through the import of status information, the related information is associated with the user id. In consequence, before users use the system, it has assigned the relevant entry instruction, and the system has stored the attribute information of users. Without active users manual input, the system can extract user related natural attributes through the user id, such as: students' grade, teachers' teaching subjects and so on, and thus solves the cold start problem in recommendation algorithm;

(2) The user interest presents the aggregation pattern, and the user group presents the aggregate form. Generally speaking, the interest of users in the same grade in the micro course system is similar basically, all developed by revolving the curriculum this grade studies;

(3) For the user interest change law, under normal circumstances, in a semester of time, the user's interest will not change very frequently; but once the new semester is coming, the user interest change will be a great degree of change. Because for one semester, the user interests are basically the knowledge points learned in this semester; but the new semester begins, the degree of attention of the knowledge in last semester will be reduced or no longer concerned;

(4) Video resource label-orientation, "Standard of Education Information Industry" issued by the Ministry of Education made provision of the specification of educational resources. It requires to make the education resources standardized, and according to the education industry standards, make the education resources labeled. That is to say, each resource will correspond to a classification label, then the extraction of the user interest will be carried out mainly based on the video tag;

(5) The score data of video resources are more objective, so the video score is more effective and efficient.

### 3.2 User interest weight calculation

Because the user behavior will be associated with the video resources directly, according to the provisions of needing to make tag-nature of education resources in "People's Republic of China education industry standard", we can get tag corresponds to the video resources in the system. As a result, according to the video resources id associated in user behavior, we can obtain the corresponding tag. The user interest weight calculation is actually the calculation of the weight of tag. The relationship among users, video resources and tag is shown in figure 2.

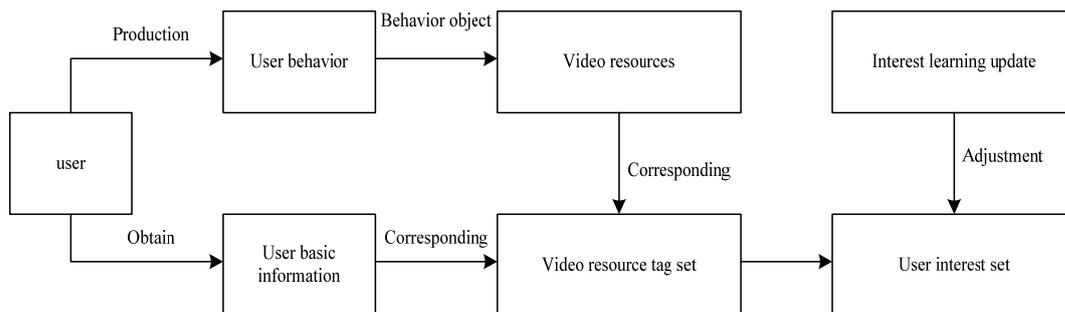


Figure 2. Analysis of the relationship among user- resources- interest

In this study, in the user weight calculation, we need to consider the following two aspects: for new users, we directly based on the basic attributes of the user, which can be directly obtained by the

system, and taken as the user's initial interest tag collection. At the same time, give the initial weights for each initial interest; for non-new users, we use the history data of the user, and through the relevant calculation, finally get the weight that each tag in the tag set users interested in corresponds to.

Based on the particularity of video resources, in the calculation of the user's interest weight, the interest weight composition is divided into five parts: the weight generated by the user for interest (video tag) search behavior, weight produced by the browsing behavior, weight generated for view behavior, weight produced for download behavior, and weight produced by grade behavior. According to the above five types of behaviors, we can get the set  $D = \{s, b, v, d, g\}$ , and at the same time, in view of five kinds of behavior responds to user interest, respectively give interest impact factor for the five behaviors, which is defined as  $\delta$ . Then, the final corresponding user interest weight calculation formula in allusion to a certain interest  $j$  is defined as:

$$W_j = \sum_{x \in D} W_{jx} \times \delta_x \quad (1)$$

In (1), the weight of user in allusion to a certain behavior happened to tag  $j$ , namely  $w_{jx}$  in (1). This paper adopts TF-IDF method to calculate the weights that the behavior type corresponds to, which is generated by the user interested tag. The original calculation formula of TF-IDF is shown as follows:

$$w_{ij} = \frac{tf_{ij} * \log(N/n_i + 0.01)}{\sqrt{\sum_{ij \in D_j} [tf_{ij} * \log(N/n_i + 0.01)]^2}} \quad (2)$$

In (2),  $w_{ij}$  refers to the weight of the feature item  $t_i$  in the text  $D_j$ ,  $tf_{ij}$  suggests the frequency of the feature item  $t_i$  in the current document,  $N$  is the number of documents contained in the training items collection,  $n_i$  is the total number of documents of the feature item  $t_i$  contained in the training item set. Thus, in the micro course system, the formula for calculating  $w_{jx}$  is defined as:

$$w_{jx} = \frac{tf_{jx} * \log(N/n_j + 0.01)}{\sqrt{\sum_{ij \in D_i} [tf_{jx} * \log(N/n_j + 0.01)]^2}} \quad (3)$$

In (3),  $tf_{jx}$  is in the current user behavior record, the frequency of  $x(x \in D)$  behavior type happened according to tag  $j$ ,  $N$  is the user behavior record of users with  $x$  behavior types in the same day in all users,  $n_j$  is the user behavior record for users with  $x$  behavior of tag  $j$  in all users.

In (3), the range of  $w_{jx}$  is  $[0, 1]$ . From the above formula, it can be found that, the value calculated by the formula can reflect the influence degree of tag  $j$  on the user interest. Because the calculation value is 1, then it suggests the tag has obvious representation and graduation in calculating the user interest, and conversely, if this calculated value is relatively small, the tag is very common and that distinction is very small.

At the same time, in order to determine the numerical factor affecting the interest in five classes of user behavior, this paper designs a questionnaire survey, to investigate the interest degree of the video resources related to tag in users searching, browsing, watching, downloading, scoring and so on behaviors. Based on the sampling investigation of 40 individuals, we can obtain the value of  $\delta_s$ ,  $\delta_b$ ,  $\delta_v$ ,  $\delta_d$ , and  $\delta_g$  are 0.1, 0.15, 0.2, 0.25 and 0.3. In view of time and environmental factors, this sampling data is small, so the questionnaire results may generate certain deviation. As a result, in the subsequent, we decided to use a certain mathematical model to make its modeling, and thus get more exact numerical.

### 3.3 User interest model representation

In the user interest model representation method, we introduce several common model representation methods, in which vector space model representation is the most mature and the most efficient and effective method. Moreover, the vector space model is consistent with the actual situation of this paper, which can be used to summarize the needs and preferences of users. Therefore, the vector space model is used to represent the output of interest model in this paper.

In this paper, the VSM expression is used to construct and express the user interest model. Each data item in the model is composed of two indexes: the user interest item  $tag_i$  and the weight value  $w_i$  corresponding to the interest item. Among them, the weight value can reflect the user's attention to the interest item.

By the calculation of (1), we get the user interest item and the weight information related to it, which in the model can be expressed as  $tag_j - W_j$ , similar to the key pairs, so the user interest model VSM method is represented into the following form :  $\{(tag_1, w_1), (tag_2, w_2), \dots, (tag_j, w_j), \dots, (tag_n, w_n)\}$   $tag_j$  ( $j \in \{1,2,3,\dots,n\}$ ) is the interest item users concerned extracted based on the user's explicit and implicit features,  $w_j$  ( $j \in \{1,2,3,\dots,n\}$ ) corresponds to the user's attention weight to  $tag_j$ .

#### 4. Collaborative filtering TOP-N algorithm based on users

For the implementation of the recommendation functions in the recommendation system, it is usually done in three steps, as shown in figure 3.

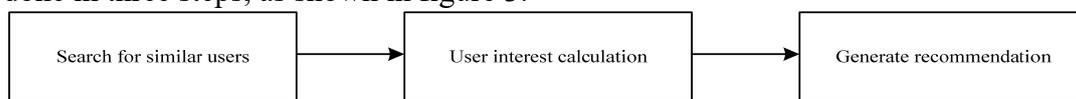


Figure 3. Collaborative filtering algorithm based on users

##### 4.1 User similarity calculation

In this paper, the user score vector is very convenient to calculate. We can calculate the angle cosine between vectors, or use the Pearson correlation coefficient. The calculation formula of the two are as follows:

$$\text{sim}(u_1, u_2) = \cos(a, b) = \frac{a \cdot b}{\|a\| \times \|b\|} = \frac{\sum_i^m r_{1i} r_{2i}}{\sqrt{\sum_i^m r_{1i}^2} \sqrt{\sum_i^m r_{2i}^2}} \tag{4}$$

$$\text{sim}(u_1, u_2) = \frac{\sum_i^m (r_{1i} - \bar{r}_1)(r_{2i} - \bar{r}_2)}{\sqrt{\sum_i^m (r_{1i} - \bar{r}_1)^2} \sqrt{\sum_i^m (r_{2i} - \bar{r}_2)^2}} \tag{5}$$

In (5),  $a$  and  $b$  respectively refers to the score vector of user  $u_1$  and user  $u_2$ ,  $r_{1i}$  and  $r_{2i}$  indicates the evaluation score of the user  $u_1$  and user  $u_2$  for video  $i$ . In (5), the calculation method of Pearson correlation coefficient compensates for the defect that the user's score vector space distance cannot be considered in cosine similarity computation. This study is based on the use of vector space model as the basis for the calculation, at the same time combined with User-CF, so it is called VSM User CF. The cosine similarity calculation method is used to calculate the similarity between users. For the user  $u$  and user  $v$ , the similarity between the two is calculated. Then the similarity formula is:

$$\text{sim}_{uv} = \frac{|\text{tag}(u) \cap \text{tag}(v)|}{\sqrt{|\text{tag}(u)| |\text{tag}(v)|}} \tag{6}$$

In (6),  $\text{tag}(u)$  and  $\text{tag}(v)$  represent the interest set of user  $u$  and user  $v$ .

In the following, the user interest data examples are given:

$$\text{Tag}(u) = \{(A, w_a), (B, w_b), (C, w_c), (D, w_d)\}$$

$$\text{Tag}(v) = \{(A, w_a), (B, w_b), (C, w_c), (E, w_e)\}$$

Then make use of (4) to calculate the similarity between the two parts:

$$\text{sim}_{uv} = \frac{|\{A, B, C, D\} \cap \text{tag}(A, B, D, E)|}{\sqrt{|\{A, B, C, D\}| |\{A, B, D, E\}|}} = \frac{3}{4}$$

The time complexity of the algorithm is  $O(|U| * |V|)$ , then it can be found that the time complexity of the algorithm is proportional to the calculated number of users, and with the increase of the number of users, the time complexity will show exponential growth. At the same time, there is little or no association between the interest of many users in the system, namely  $I(u) \& I(v) = 0$ . In this case, the

use of this method in the calculation, whether it is in the calculation of time, or in the system overhead, will bring a lot of burdens. In order to solve this problem, in this paper, in the calculation by the above formula, we will make a simple point conversion. By adding an intermediate data, the data contains a interest (i.e., tag of video resources) and the user list with the interest, in which all the information is correspondingly stored in the way of the key pairs, represented as  $\text{tag} - \text{UserList}: \langle \text{tag}, \text{userlist} \rangle$ .

Through the reversion of the calculation method, then the calculation of user  $u$  similarity user set can be concluded as the following steps:

- (1) To obtain the user  $u$  interest set  $I(u)$ , and all tag in  $I(u)$  is taglist;
- (2) To increase  $\forall \text{tag} \in \text{tagList}, \langle \text{tag}, \text{userlist} \rangle$  to the relevant users  $S(u, K)$  of user  $u$ ;
- (3) For  $\forall v \in S(u, K)$ , to calculate the similarity  $S(u, K)$  with the user  $u$  by using the cosine similarity method and form a similar user set of user  $u$ .

#### 4.2 User interest calculation

The target user nearest neighbor set is  $S(u, K)$ , and make use of the similarity between each user and the target user in the collection to predict the interest of the target user to the video without effective behavior. The prediction formula is shown as follows:

$$p(u, i) = \sum_v \text{sim}_{uv} b_{vj}, \quad v \in S(u, K) \cap U(i) \quad (7)$$

In (7),  $S(u, K)$  suggests  $K$  nearest neighbor sets with user  $u$ ,  $U(i)$  indicates the user set interested in  $\text{tag}_j$ , and user  $v$  is the user who is interested in  $\text{tag}_j$  in the nearest neighbor of user  $u$ .  $\text{sim}_{uv}$  is the interest similarity between user  $u$  and user  $v$ , and  $b_{vj}$  represents the interest weight of user  $v$  to  $\text{tag}_j$ .

Eventually, we select  $M$  interests with the largest interest value in the user interest set, and we define it as  $I_m = \{I_1, I_2, \dots, I_m\}$ .

#### 4.3 Recommendation generation

According to the above user interest set  $I_m$  we get, we access to the video resources set related to the user interest, defined as  $V_n$ . Before the recommended results are presented to the user, we first of all make a descending order of  $V_n$  video resources, because from the Internet users usual use habits, we know that the more front the ranking resources, the higher the degree of attention of users.

In the calculation of ranking weights, we consider two factors: the user interest value and video evaluation. We adopted the two considerations, and finally formed the ranking weights of the video. The display position of the weights and the video on the page is proportionally related. The greater the video weight, the more forward the ranking. The weight calculation formula is as follows:

$$w_v = p(u, i) \times \alpha + R \times \beta, \quad \text{among } \alpha + \beta = 1 \quad (8)$$

In (8),  $p(u, i)$  is the results calculated in (5), suggesting the user preference to the video with  $\text{tag}_j$ , and  $\alpha$  refers to the effect factor of user preference to the video in the ranking weight. The  $\alpha$  value is set to 0.75,  $R$  is the video score, and  $\beta$  is the effect factor of the video score on the ranking weights. By (6), we finally obtained the ranking weights of the video. The paper, at last, according to the video ranking weight, recommends the video after ranking to the user, and what is finally presented to the user will be the video resources that users are interested in with high evaluation.

### 5. Conclusion

The work done in this paper is as follows: (1) the user interest model is established, which is represented by the vector space model with rather wide adaptability. At the same time, the data update mechanism of the model is established. According to the feedback generated by the user behavior, the user's interest in the interest model is enhanced or weakened. (2) the selection and application of the recommendation algorithm, according to the practical application background of the system, the user based collaborative filtering recommendation algorithm is applied. The experimental results show that the algorithm has a good practical effect in this system.

## References

- [1] Hloušek, M. (2014). Zero lower bound on interest rate: application of DSGE model on Czech economy. Palacký University.
- [2] Nowak, P., & Romaniuk, M. (2014). Application of the one-factor cir interest rate model to catastrophe bond pricing under uncertainty. *Journal of Automation Mobile Robotics & Intelligent Systems*, 8(3), 19-27.
- [3] Mingqian, L. I. (2014). Insurance calculation of bankruptcy probability of constant interest rate model under dependent negative. *Guide of Science & Education*.
- [4] Sidorov, G., Gelbukh, A., Gómezadorno, H., & Pinto, D. (2014). Soft similarity and soft cosine measure: similarity of features in vector space model. *Computación Y Sistemas*, 18(3), 491-504.
- [5] Ghosh Dastidar, A. (2016). Personalized bookmarking of textsite applications via a text message.
- [6] Littleton, J., & Zieglgänsberger, W. (2015). From details to large scale: the representation of environmental positions follows a granularity gradient along the human hippocampal and entorhinal anterior-posterior axis. *Hippocampus*, 25(1), 119.
- [7] Bayo Moriones, J. A., Galdon-Sanchez, J. E., & Martinez-De-Morentin, S. (2016). The process of wage adjustment: an analysis using establishment-level data. *Economic and Industrial Democracy: An International Journal*(2).
- [8] An, D., Zheng, X., Rong, C., Kechadi, T., & Chen, C. C. (2015). Gaussian Mixture Model Based Interest Prediction In Social Networks. *IEEE, International Conference on Cloud Computing Technology and Science* (pp.196-201). IEEE.