Research and Application on Personalized Recommendation of Potential Friends and Books Base on Collaborative Filtering

He Chengdong and Shu Chang

1 Lian Yi Technology Co., Ltd., Guangzhou 510000, Guangdong, China

*Corresponding author’s e-mail address: heironhand@163.com

Abstract. The personalized recommendation of potential friends and book resources is important for improving the personalized service of digital campus. In this context, this paper proposes a recommendation method of potential friends and book classes based on collaborative filtering. The establishment of users profiles and the implicit rating prediction model by mining user demographic attributes and book borrowing features. It can improve social environment and utilization of library resources base on this method. The result shows that the model can effectively improve the quality of personalized service recommendation of potential friends and book resources.

1. Introduction

Construction of intelligence service and sharing platform break through the "Isolated Data Island". The platform integrates data of campus card, library borrowing, student achievement, attendance, entrance guard and network access log in a big data environment[1]-[3]. The key is mining and analyzing the rule of characteristic data of school students and managers in digital campus and smart campus.

One of the goals of the construction of digital campus platform and smart campus is to provide personalized services for students, and personalized recommendation of potential friends and book resources is an important part of personalized service[4][5]. Recommendation technology can be more intelligent to recommend the potential friends and learning resources to users with similar behaviour feature, so as to promote the development of the personalized service on campus.

2. Existing problems

At present, there exists few researches on recommending service for campus big data platform[6]. It does not form a deep mining for the campus users' basic attributes and real behavior data, and builds a model for interest prediction[7][8]. Therefore, recommendation technology applied to campus scene can alleviate information overloading, improve communication efficiency, and form personalized service. The method is to provide recommendation information based on users' basic attributes and interest preferences, and it also can effectively guide users to make the right decisions. The work we need to do, in this paper, is to establish a recommendation model for potential friends and book resources on campus, so as to improve the social interaction efficiency and the utilization rate of library resources.
Personalized recommendation technology usually adopts collaborative filtering algorithm that obtaining the user's eigenvector, calculating the similarity between users profiles, and recommend and predict[9][10]. Recommendation technology research mainly focuses on business service, but there are few researches on recommendation of digital campus resources. Therefore, it is necessary to research collaborative filtering technology applied to campus friends and Book Resources recommendation.

This paper proposes some solutions to the issues discussed above. Mining students' basic attributes and borrowing data, establishing user interest feature models, and using collaborative filtering algorithm to recommend friends and books. It is beneficial to improve the personalized service of the campus.

3. Existing problems

3.1. Student demographic attributes
The demographic attributes of the students are Student ID, Name, Sex, Birth date, ID, Native place, Department code, Professional code, class ID, role ID, Cell-phone number, E-mail, Account status, and School status in the management system of College Students' behavior analysis. This paper quantifies four factors: student ID(Grade), gender, ID card (Regions), department code (Department).

- Student ID(Grade): The first two numbers in the school number represent grade. Two methods can be used to quantify grade, for example, the sparse representation method is expressed as (1000), 15 is (0100), 16 is (0010), and 17 is (0001). Another method of quantification is (1, 2, 3, 4), respectively.
- Sex: 'male' is quantified to be '1', 'female' is '0'.
- ID card(Regions): The first two numbers represent regions. The representation method is consistent with the Student ID(Grade).
- Department code (Department): The first two numbers represent regions. The representation method is consistent with the Student ID(Grade).

The table 1 indicates that the four feature quantization results are normalized and form user profiles matrix.

| F1 | F2 | F3 | F4 |
|----|----|----|----|
| U1 | H11| H12| H13| H14|
| U2 | H21| H22| H23| H24|
| ...| ...| ...| ...| ...|
| U1 | H11| H12| H13| H14|
| ...| ...| ...| ...| ...|
| Un | Hn1| Hn2| Hn3| Hn4|

The feature vectors of user n and user m are defined as Hn=<Hn1, Hn2, Hn3, Hn4>, Hm=<Hm1, Hm2, Hm3, Hm4>, respectively.

Pearson similarity method is used to calculate the similarity of user demographic characteristics.

\[
sim(U_i, U_j) = \frac{\sum (H_{ik} - \overline{H_i})*(H_{jk} - \overline{H_j})}{\sum (H_{ik} - \overline{H_i})^2*(H_{jk} - \overline{H_j})^2}
\]

\(H_{ik}, H_{jk}\) is the quantized values of user n and m corresponding to the feature k respectively. \(\overline{H_i}, \overline{H_j}\) is average value of all the features by the user n and m.

3.2. User profiles
The library borrowing information includes Book ID, Borrower ID, Student ID, Staff ID, Barcode, Book name, Author, Press, Publication date, Library ID, Borrow date, and Return date in the management system of College Students' behaviour analysis.
Collaborative filtering usually mining users' potential interests, through analysing all rating data. But the books is special, they are clustered through the Chinese Library Classification that clusters books into n categories, and table 2 shows the form of book classification in Chinese Library Classification. Different classes of book can respond to interest of student. The cluster frequency borrowing is used to be converted to rating of items when lack of book rating and frequency borrowing of one book is one for a single user.

Table 2. Chinese Library Classification

| Class | Name | Class | Name |
|-------|------|-------|------|
| A     | Marx doctrine, Lenin doctrine, Mao Zedong thought and Deng Xiaoping theory | T-9 | Industrial economy |
| B     | Philosophy, Religion | TB | General industrial technology |
| C     | social science general introduction | TD | Mineral engineering |
| D     | Politics, Law | TE | Oil and gas industry |
| E     | Military | TF | Metallurgical industry |
| F     | Economics | TG | Metallization, metal technology |
| G     | Culture, Science, Education, Sports | TH | Machinery, instrument industry |
| H     | Language, Writing | TJ | Weapons industry |
| I     | Literature | TK | Power industry |
| J     | Art | TL | Atomic Energy Technology |
| K     | History, Geography | TM | Electrician technology |
| N     | Natural science introduction | TN | Radio electronics, Telecommunications Technology |
| O     | Mathematical science and chemistry | TP | Automation technology and computing technology |
| P     | Astronomy, Earth Science | TQ | Chemical industry |
| Q     | Bioscience | TS | Light industry, Handicraft industry |
| R     | Medicine, Health | TU | Architecture Science |
| S     | Agricultural Science | TV | Water conservancy engineering |
| T     | Industrial Technology | U | Transportation |
| T-0   | Industrial technology theory | V | Aeronautics, Astronautics |
| T-1   | The status and development of industrial technology | X | Environmental Science, Safety |
| T-2   | Institutions, Groups, Meetings | Z | Science |
| T-6   | Reference book |

In the context of implicit user rating in the frequency of user borrowing books, they are labeled as “not interested” when borrowing times are very few, On the other hand, it is marked as “very interested”. the higher the frequency is, the higher interest is, and tends to be stable. Therefore, in this paper, a nonlinear relation is used to indicate the frequency of books borrowing and their interest in book classes.

\[ r(x_i) = \alpha(1 - (1/e)^\mu) \]

\( r(x_i) \) is borrowing frequency, it is marked as interested if above \( \mu \) times. \( \alpha \) is a parameter to determine the scope of the rating. In order to be consistent with the scoring mechanism, This paper sets up \( \alpha = 5, \mu = 5 \).

After the above transformation, the user rating matrix is indicated by table 3.
Table 3: User profiles matrix from implicit rating

|       | C₁  | C₂  | ... | Cₖ  | ... | Cₘ  |
|-------|-----|-----|-----|-----|-----|-----|
| U₁    | r₁1 | r₁2 | ... | r₁ₖ | ... | r₁ₘ |
| U₂    | r₂₁ | r₂₂ | ... | r₂ₖ | ... | r₂ₘ |
| ...   | ... | ... | ... | ... | ... | ... |
| Uₙ    | rₙ₁ | rₙ₂ | ... | rₙₖ | ... | rₙₘ |

Pearson similarity method is used to calculate the similarity of user profiles.

\[
\text{sim}(U_i, U_j) = \frac{\sum (r_{ik} - \bar{r}_i)(r_{jk} - \bar{r}_j)}{\sqrt{\sum (r_{ik} - \bar{r}_i)^2 \cdot \sum (r_{jk} - \bar{r}_j)^2}}
\]  

(3)

Obtaining \( K \) adjacent users whom are recommended to the target user with higher similarity through comparison. Then prediction by the following formula model.

\[
P_{i,C} = \frac{\sum_{j \in N} \text{sim}(i, j) \cdot R_{j,C}}{\sum_{j \in N} \text{sim}(i, j)}
\]  

(4)

\( N \) represents an adjacent user set that has a rating on the item \( C \), \( R_{j,C} \) is rating by user \( j \).

3.3. Personalized Recommendation Process

The process of personalized recommendation on friends and book classes is structured in four steps as follows. The figure 1 shows the flowchart of friends and book classes recommendation.

- Step 1: The related data is obtained through the server, and the required data is filtered and obtained.
- Step 2: The eigenvalue of labels are analyzed to get the user profiles between students and tags.
- Step 3: Obtaining \( K \) adjacent users whom are recommended to the target user with higher similarity between user profiles.
- Step 4: Top \( N \) items are recommended for predicting the missing rating of target user’s feature.

![Figure 1. The flowchart of friends and book classes recommendation](image)

4. Experimental results and analysis

The experimental data are 100 student’s demographic information and book borrowing data. The data set that divided into training set and test set based on the ratio of 8:2 are filtered to get 1084 effective
borrowing records. The mean absolute error (MAE), widely used in collaborative filtering recommendation, is used to measure the prediction accuracy.

\[
MAE = \frac{1}{N} \sum_{i=1}^{N} |r_i - \hat{r}_i|
\]  

(5)

\(N\) is the number of test samples, \(r_i\) and \(\hat{r}_i\) is real and prediction rating respectively.

The figure 2 is the prediction accuracy of the implicit rating of the book classes. The experimental results show that when the nearest users Top \(K\) is 20, the prediction accuracy of book classes implicit rating is stable at 1. Top \(N\) book classes are recommended for target user, and users with higher similarity as potential friends.

![Figure 2. Prediction accuracy of book classes implicit rating](image)

5. Conclusion
A personalized recommendation service based on the student demographic attributes and frequency borrowing of digital campus users is realized by data mining to obtain user profiles and establish potential friends and book classes recommendation models. In this paper, a recommendation collaborative filtering method using a potential friend and book classes is practical. It will improve the efficiency of social interaction and the utilization of library resources. Meanwhile, personalized recommendation service can also be applied to other scenes of digital campus, so as to promote the construction of smart campus.

References
[1] Zheng T, Sheng X L. Intelligent Campus Personalized Information Push Service Base on User Interest [J]. Journal of Zhejiang Shuren University. 2017, 17(1):15-19+27.
[2] Ding D, Li J, Wang H, et al. Student Behavior Clustering Method Based on Campus Big Data[C]// International Conference on Computational Intelligence and Security. IEEE Computer Society, 2017:500-503.
[3] Zhang N. A Campus Big-Data Platform Architecture for Data Mining and Business Intelligence in Education Institutes[C]// International Conference on Machinery, Materials, Environment, Biotechnology and Computer. 2016.
[4] Binas A, Charlin L, Levinshtein A, et al. Indigo Books and Friends Recommendation[J]. 2009.
[5] Gao B, Du S, Li X, et al. Research on the Application of Persona in Book Recommendation System[C]// Journal of Physics Conference Series. Journal of Physics Conference Series, 2017:012023.

[6] Hu H, Long D. A Study on Key Techniques of Wisdom Campus Information Recommendation Platform Based on Big Data[C]// International Conference on Intelligent Transportation, Big Data and Smart City. IEEE, 2016:56-58.

[7] Liang Z. Research on student behavior analysis and prediction method based on campus big data [D]. Xi'an University of technology, 2017.

[8] Lindsey R V, Khajah M, Mozer M C. Automatic discovery of cognitive skills to improve the prediction of student learning[C]. Advances in neural information processing systems, 2014: 1386-1394.

[9] Liu Q W. Research on Recommendation System Based on Collaborative Filtering [D]. University of Science & Technology China, 2013.

[10] Quan Z. Collaborative filtering recommendation based on user personality[C]// International Conference on Information Management, Innovation Management and Industrial Engineering. IEEE, 2014:1041-1046.