

Research on the Application of Persona in Book Recommendation System

Gao Baozhong¹, Du Shouyan^{2,*}

1. Institute of Information Science and Technology, Shandong Normal University
2. Super-computing Center, Computer Network Information Center, Chinese Academy of Science
Jinan¹, Beijing², China

E-mail: gaobaozhong@sdsu.edu.cn, dushouyan@hotmail.com

Abstract— Currently, there still exists a host of problems in the book recommendation system, such as low accuracy, weak correlation and poor pertinence. Aiming to unravel these problems, this paper based on the theory of big data and data mining technology, through analyzing internet user behavior and the "5C" model of personal credit evaluation, combined with joint impact weight calculation method, which involves user grade, borrowing credit, book friend recommendation degree, book friend recommended adoption degree, borrowing frequency, borrowing number, and borrowing time interval. User activity and credit are also taken into account in the process of establishing user tagging system so as to build classified book recommendation service. This method is of universal meaning to the book recommendation service of smart campus with user as the core under big data environment.

Keywords- *persona; internet credit; data mining; smart book recommendation system; big data*

I. INTRODUCTION

Persona is first proposed by the father of interactive design Alan Cooper, he believes that persona is the virtual symbol of real user, which is a target user model based on a series of real data. According to the combination of its user attributes and behavioral data and the combination of dynamic tags and multi-level tags, it can better identify the user's community characteristics and thus customize service for the user. The intelligent recommendation service based on the "persona" has obtained a wide range of applications in the Amazon, Ali, Jingdong and other electricity and information services business^[1].

Li Bing explored the characteristics of cigarette retailers based on big data technology and K-means algorithm, realizing the intelligent recommendation for the order of retailers. Wu Mingli and Yang Shuangliang analyze the behavior data of mobile users, combining the two dimensions of time and space and with the use of spark cluster, solving the problems of push service accuracy and large amount of user data^[2]. In the space, the book recommendation areas focus on different from other areas of persona, so it's need for their own characteristics of the corresponding design of the persona, in order to achieve accurate recommendation^[3]. In this paper, we use spyder integrated development environment and the data of user behavior real-time transmitted to the system data users, through the user multi-dimensional comprehensive analysis to realize the establishment of integral values, and constantly

optimize and update the persona of system, in order to achieve the content pushed to the user meet the highest degree of customer demand and interest preferences of the content, and stimulates the user's interest in reading.

II. BIG DATA TECHNOLOGY AND CREDIT EVALUATION

A. Big Data Technology

At present, "big data" does not form a recognized definition. National Institute of Standards and Technology NIST's defines big data as data capacity, data acquisition speed or data representation limits the data analysis and processing capabilities of using traditional relational methods, thus need to use the mechanism of horizontal expansion to improve processing efficiency^[4]. It is now generally considered that "big data" contains massive structured, semi-structured and unstructured data, when compared with traditional relational databases^[5], big data-related techniques have a greater advantage in processing speed and storage capacity when dealing with semi-structured and unstructured data. Common big data processing models usually include data acquisition, data preprocessing, data and text mining and data visualization four modules^[6].

B. Credit Analysis and Rating

Personal credit investigation, refers to the credit and related institutions based on the agreement to collect, process, and store information scattered in different institutions about the user credit, and the formation of credit data provide a reasonable basis of the user's credit situation for a business activity. Presently, the international financial institutions assess customer credit mainly use the "5C" model expert analysis method to do risk analysis for users, accessing user's ability to repay and willing according to the character, capacity, capital, collateral, and condition five aspects, which can more accurately reflect the user's credit performance in a certain way to determine the degree of risk^[7]. In the application of the library book recommendation system based on the persona, we make use of the data mining technology to obtain the user's data from the intelligent campus system and the library system, analyzing the subjective and objective factors that affect the user's credit level, then, the user rating model quantifies the influencing factors with specific scores and synthesizes their performance to form data feedback, thus building a more accurate multi-dimensional persona.

III. BUILD USER RATING SYSTEM

A. Data Collection

Data is the core of the user's interest tag. This project uses data mining technology to collect the relevant data of user from library and intelligent campus system. The data source of persona can be divided into explicit data and hidden data two aspects. As shown in Table 1. The explicit data includes user attributes and library book data, which are relatively stable. For this kind of information, recommendation system does more work as data cleansing without the need of modeling predictions. The hidden data is divided into two categories, namely, dynamic user data and platform data, which are hidden in the user random internet behavior, and need to mine hidden information and extract the information ontology through data analysis and other techniques.

TABLE I. SYSTEM DATA INFORMATION

data	type	description
Explicit data	User attributes	Including the user (graduate, teacher, undergraduate) ID number, name, gender code, department code.
	Library Book Data	Including the library book barcode, book name, reader barcode, borrow date, return date, renewal logo.
Hidden data	Dynamic user data	Including the user search book records, collection of book records, borrowed records.
	Platform data	Data obtained from the user and the platform for book interaction.

B. User Behavior Image Dimension

According to the principle of presenting similarity of the characteristics within the user group, showing difference between the user groups to achieve the users grouped. This paper analyzes the user behavior, based on the collected and mining data, from different dimensions adding hierarchical organization of the tag for user, in order to make the establishment of the user's tag library, thus achieving the user's behavior, attribute tag. The persona based on our smart campus users mainly include as shown in Table 2.

TABLE II. PERSONA DIMENSION

type	description
Basic information	Including the users (graduate students, teachers, undergraduates) ID number, name, age, gender, faculties, hobbies, etc.
Borrowing habits	Including the user's borrowing period, borrowing frequency, borrowing time, borrowing time interval, the influencing factors of borrowing books and credit, etc.
Access preferences	According to the number of user access different types of books in the network to determine the user preferences, this persona is portrayed on a monthly basis.
Search Information	Describe the condition of user using the library system search engine to retrieve the content, which can more accurately reflect user's points of interest and needs than the access preferences.
Other features	Including the user active situation, user professional classification, etc.

C. User Credit Evaluation Index

This paper is based on the dimension of 5C method when designing the index system of user credit evaluation in colleges, through the basic information, borrowing habits, accessing preferences, search information and other characteristics five dimensions, combining with the service of book recommendation system and obtained data for user credit evaluation, and the credit points determine how much the recommended function and the number and frequency of books the user can obtain. Index selection basis and score description of impact are shown as follows:

- User grade: Low grade users due to have less courses, and more adequate spare time, thus having rather high willingness to borrow. With the increase in user grade, the borrowing willingness becoming lower to some extent, but the system will increase the proportion of recommended professional or competition books.
- Borrow credit: Determined by the user whether is scheduled to return the borrowed books, and the longer the period expires, the lower credit score user has.
- Book friends recommended degree: In book recommendation system, the user with similar interests can recommend books to other users after reading some feeling good, and the higher the recommended number of users, the higher the corresponding score.
- Book friends recommended adoption degree: On the basis of mutual recommendation between book friends, combining with the number of other users who adopted the books to determine, and the higher number of users adopt that book, the higher scores the recommended users will enjoy.
- Borrowing frequency: The higher frequency of borrowing users on behalf of the higher demand for borrowing.
- Borrowing number: The number of borrow book is an indirect reflection of the user's borrowing needs, the more books borrowed, the higher user borrowing needs.
- Borrowing time interval: For some users, they no longer borrow books for a long time after one time, whereas some other users borrow time interval represent as coherent cyclical, showing a higher borrowing demand.

D. User Characterization

User characterization is user tagging. The tag is a feature of highly refined hierarchical catalog form which is analyzed by user information. Based on the user's behavior data and a certain model algorithm to get the user's tag weight, combining with different dimensions of tag to form a scoring system to distinguish recommended books priority between users with different characteristics. The static ontology is relatively stable and the weight is set to 1. In this paper, the weight of persona tag is mainly considered by the user grade, borrowing credit, book friend recommendation degree, book

friend recommended adoption degree, borrowing frequency, borrowing number, and borrowing time interval seven factors, and each tag weight calculates as following:

1) User grade

User grade belongs to static ontology, but is need to combine this factor in the book recommendation system to recommend, and different grades of users book borrowing demand is also different. The user's grade on the weight of the body is represented as $G(label)$, which is the length of time of user until graduation divided by the total length of four years, namely:

$$G(label) = \frac{T_{end} - T_i}{T_{end} - T_{start}} \quad (1)$$

2) Borrowing credit

The user's borrowing credit mainly about after the user borrow books, whether is return the book on time, and borrowing credit on the weight of the body is represented as $C(label)$, which is the length of time of user borrow books in system divided by the length of time until user return the book, namely:

$$C(label) = \frac{T_{bsystem-end} - T_{bsystem-start}}{T_{b-end} - T_{b-start}} \quad (2)$$

3) Book friend recommendation degree

The weight of the book friend's recommendation on the ontology $R(label)$ refers to in the mutual recommendation module of the book recommendation system, the ratio of the number of books recommended by the user to other users to the total number of recommended books at the same period, namely:

$$R(label) = \frac{R_{label}}{\sum_{i=0}^n R_i} \quad (3)$$

4) Book friend recommended adoption degree

The recommendation of the degree of adoption on the ontology $A(label)$ refers to in the mutual recommendation module of the book recommendation system, the proportion of the number of adopted books recommended by the user to other users divided by the total value of adopted recommended books at the same period, namely:

$$A(label) = \frac{A_{label}}{\sum_{i=0}^n A_i} \quad (4)$$

5) Borrowing frequency

The borrowing frequency of user is used to describe the number of times of borrowing in unit time, which is portrayed on a monthly basis. Borrowing frequency on the weight of the body is represented as $F(label)$, which is the ratio of the number of times the user borrows the book to the total number of borrowing in the system at the same period, namely:

$$F(label) = \frac{F_{label}}{\sum_{i=0}^n F_i} \quad (5)$$

6) Borrowing number

According to the user's borrowing in the book recommendation system to do the borrowing list, in months

as a unit, and the borrowing number of ontology $N(label)$ is the ratio of the number of users borrowed to the total number of borrowed in system at the same period, namely:

$$N(label) = \frac{N_{label}}{\sum_{i=0}^n N_i} \quad (6)$$

7) User average borrowing time interval

The average borrowing time interval $U(label)$ indicates the sum of time differences between the user's borrowing actions divided by the number of users borrowing except the same borrowing date, which is portrayed on a daily basis. The larger weight value represents longer time interval, m is the number of users borrowing except the same date, and the formula is:

$$U(label) = \frac{\sum_{i=0}^n (U_{i-start} - U_{i-lastend})}{m} \quad (7)$$

E. User Rating Model

In this paper, the multi-dimensional user tagging system based on the data information obtained from intelligent campus and library system can provide detailed data reference for user's rating evaluation. The persona score model is shown as below:

$$people = \sum_{i=1}^n label_i p(i) \quad (8)$$

$$= \sum_{i=1}^n label_i \{ [G(label) + C(label) + R(label) + A(label) + F(label) + N(label) + U(label)] \}$$

$$= \sum_{i=1}^n label_i \left\{ \left[\frac{T_{end} - T_i}{T_{end} - T_{start}} + \frac{T_{bsystem-end} - T_{bsystem-start}}{T_{b-end} - T_{b-start}} + \frac{R_{label}}{\sum_{i=0}^n R_i} + \frac{A_{label}}{\sum_{i=0}^n A_i} + \frac{F_{label}}{\sum_{i=0}^n F_i} + \frac{N_{label}}{\sum_{i=0}^n N_i} + \frac{\sum_{i=0}^n (U_{i-start} - U_{i-lastend})}{m} \right] \right\}$$

$label_i$ indicates the information tag of the user on a dimension, $p(i)$ represents the weight of the tag, $G(label)$ 、 $C(label)$ 、 $R(label)$ 、 $A(label)$ 、 $F(label)$ 、 $N(label)$ 、 $U(label)$ are corresponding to user grade, borrowing credit, book friend recommendation degree, book friend recommended adoption degree, borrowing frequency, borrowing number, borrowing time interval seven factors on the weight of the tag, m is the number of users borrowing except the same date. According to the weighting formula of the user's score model, the weight value of different labels on the ontology of each user is calculated, then the normalized score of the user is obtained and the result is stored in the MySQL database.

F. Persona Management

Using K-means clustering method to carry out persona management, combining with the statistical methods of cluster analysis, which can better explain and describe the user's behavioral characteristics. The goal of clustering is to synthesize the data into different classes or clusters according to the similarity of data without any prior knowledge, so that the elements within the cluster have high similarity, while the elements similarity between the cluster

is low, so it's also known as non-supervised classification [8]. The clustering process is as follows:

1. The data is normalized to a range of [0,1] through linear function conversion method.
2. Set the cluster center point $k = 8$ to create the initial partition, that is, selecting 8 elements from unclassified collection U in random, as the center of eight clusters respectively.
3. The similarity of the remaining elements to 8 clusters is calculated separately, and the elements were classified into clusters with the highest similarity.
4. According to the clustering results of the previous round, the center of each eight clusters is recalculated, and the calculation method is to calculate the arithmetic mean of the respective dimensions of all the elements in the cluster.
5. Remove all elements of the new cluster center from the collection, then repeat the third step of the method and re-clustering.
6. Repeat steps 4 and 5 until dissimilarity between the results of this and previous round of clustering are less than the set threshold.
7. Output the final clustering results.

It adopts sliding window filtering algorithm with incremental update mechanism, allowing to achieve in the update operation, only updating the places that needs to be changed and will not repeated. It is required to calculate the user data and store the intermediate values on a monthly basis, and obtain the new persona in the way of association.

IV. EXPERIMENTAL RESULTS AND ANALYSIS

A. The Presentation of Persona

According to the weight formula, the user's information tag normalized to a range of [0,1], combining with the persona mathematical model for visualization processing, which can intuitively show the weight of user's priority in the recommendation system rated by user's different behavior. Figure 1 shows the corresponding persona weight value for the random selection of 200 users in the borrowing time interval and borrowing number two dimensions.

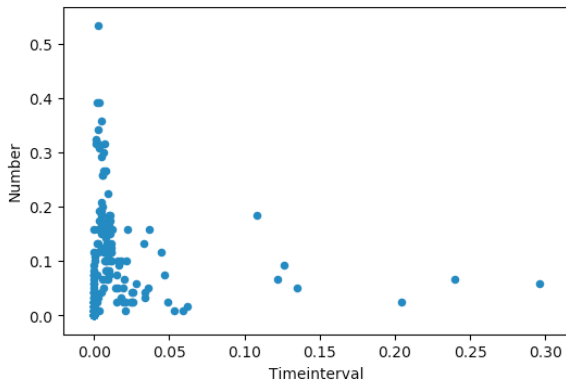


Figure 1. User borrowing time interval and user borrowing number Two-dimensional persona

According to the number of user scores, prioritizing core users, general users, and inactive users, evaluating the

potential recommended value for each group. The user is divided into 24 groups as shown in table 3. According to the scoring system, it's divided into one to five level users in the range of [0,1] with a scale interval of 0.2, corresponding to the user's overall demand, activity and credit rating.

TABLE III. USER GROUP CLASSIFICATION

User grouping			
Travel preferences class	Application preferences class	Financial preferences class	Anime class
Medical class	Education preferences class	History preferences class	Technology preferences class
Sports preferences class	Photography class	Legal preference class	Social science class
Digital preferences class	Gourmet preferences class	Environmental preferences class	Exam preferences class
Film and television preferences class	Game preferences class	Music preferences class	Car preferences class
Healthy health class	Fashion preferences class	Building preferences class	Electronic machinery class

B. Application Implementation Function

The recommendation system is a collection of software tools and technologies that provides project recommendation information for users [9]. Based on the persona, the book recommendation system according to user's behavior and borrowing records to recommend, and dynamic update content with user's interest changes. The specific application functions are as follows:

1) *Monthly recommendation*: According to the user's rating level to determine the recommended number, recommending more books to the active (high demand) level users, while recommending relatively smaller number of books to the lower level (low demand) users. On a monthly basis, the system combining with the user's preferences and real-time book list gives a comprehensive recommendation. For new users who do not borrow books, the system refers to the user's college information and recommends the most popular books recently to them, in order to solve the new user's cold start problem.

2) *Time-based recommendation*: According to the user's basic behavior persona, at the beginning of the semester, the end of the semester and before the holiday the recommended system can based on the previous grade professional books borrowing list at the same period, and appropriate increasing in the proportion of recommended books; The system in accordance with the real-time mining of user interest preferences to recommend accordingly.

3) *Book friends recommend*: The users in the book recommendation system can recommend books to each other, and the attitude to books of recommended users can

be used as one of the evaluation criteria for user interest preferences.

4) *Association recommendation*: Based on data mining technology, through the data association and correlation between the data mined from the collection, analyzing the user's borrowing behaviors, so as to achieve the association recommendation.

5) *Others borrowed recommendation*: Users with similar preferences can recommend each other's preferred books, making it easier for users to find their own interest books or groups.

V. CONCLUSION

The traditional library service fails to meet the demands of readers, while resource digitization as well as the establishment of user-based digital library have become development trends. Persona-based recommendation service through making analysis of user behavior, taking the initiative to provide users the recommended information they may need. One of the application is DM-based persona technique in library book recommendation system with bright prospect. It truly realizes user-centered design and creates larger value. In this thesis, the accurate personal credit scoring model established via data mining, requires multi-dimensional information for the sake of giving a scientific weight to each index. Thence the ultimate goal to truly reflect the credit level of the evaluation object will be accomplished. This method provides a universal reference for user-based digital library book recommendation service of the smart campus under big data environment.

REFERENCES

- [1] Jiang Jianwu, Li Jingwen, Lu Yanling. Intelligent information push method based on user portrait [J]. Microcomputer & its Applications, 2016, 35 (23): 86-89, 92.
- [2] Wu Mingli, Yang Shuangliang. Study and application of user portrait in content push [J]. Computer Knowledge and Technology, 2016, 12 (32): 255-259.
- [3] Niu Wenjia, Liu Jiqiang. User network behavior portrait [M]. Beijing: Electronic Industry Press, 2016: 3-51.
- [4] Cooper M, Mell P. Tackling Big Data. NIST, 2012.
- [5] Li Bing, Wang Yue, Liu Yongxiang. Application of k-means based persona and intelligent recommendation in big data environment [J]. Modern Computer, 2016, 16: 11-15.
- [6] Ren Juan. Reader's behavior mining and analysis of big data technology and application [J]. Editor of the Friends, 2017, 1: 31-34.
- [7] Wang Guan. Based on the user's Internet behavior data personal credit evaluation system construction analysis - to sesame credit as an example. Beijing: Beijing Jiaotong University [D], 2015.
- [8] Chen Li. Research on clustering algorithm in data mining [D]. 2007.
- [9] Ricci F, Rokach L, Shapira B. Recommender Systems Handbook[M]. USA: Springer, 2011:5-6.