Research on Personalized Recommendation Based on User Implicit Preference

Hai-he TANG
Rizhao Polytechnic, Rizhao 276826, China
36079285@qq.com

ABSTRACT: As the rapid development of the Internet and the e-commerce technology, more and more online shop based on e-commerce technology need a personalized recommendation technology to help them achieve their information filtering and automatic recommendation of goods. Therefore, how to analyze and mining these data, and understand the customer's preferences and preferred mode, then designed personalized websites to meet the needs of different customer groups, thereby increasing competitiveness, it has become imperative. On account of the deficiencies of traditional collaborative filtering, this paper uses a collaborative filtering based on preferences of user recommendation algorithm by comparison. Using Server log analyzes users how long on the page and how many times visiting the page, to find users are prefer to some product, so that we can make the relevant recommendation.

1. Introduction
As the pace of life increases, many users do not necessarily grade the products they are interested in when browsing the website. This makes it difficult to understand the user's interests and thus recommend products to them. For large business websites, the number of products and users is very large, and the user grading products are very little, so it is difficult to analyze the user's interest through user ratings.

Although the user's interest cannot be clearly obtained by scoring, the user's general interest can be exposed based on the number of times the user visits the website and the time spent on the webpage. This paper proposes a design method of implicit feedback personalized recommendation system. The method uses the user's potential implicit preference instead of the explicit user score, and uses the server-side log to analyze the length of time the user stays on the page he visits and the number of times a page is accessed, and discovers the potential of the user for a certain type of product and make relevant recommendations.

2. Research on personalized recommendation based on user implicit preference

2.1 User potential preferences
Generally speaking, when a user stays on a certain page or when the frequency of clicking a page is high, the user has a higher potential preference for the content. It is thus possible to implement recommendations for users in accordance with this aspect. A personalized recommendation system consists of two core parts, namely the judgment of user interest preferences and the implementation of automatic recommendation output. The user personalization model is established to provide users with fast and accurate referral services. User interest is calculated by analyzing the
user's browsing behavior, for example, browsing time and number of clicks, to obtain their interest preference information \cite{4}\cite{5}. The automatic recommendation part clusters users with similar preferences, finds the most frequently viewed information items in each cluster group, and ranks the top-N information items according to the user's satisfaction. After the information processing, to provide users with referral services. In this way, by analyzing the implicit information feedback in the user's browsing behavior, the user's actual interest in the viewed item can be judged \cite{6}.

This method requires the user to be a registered user of the system. After the user logs in to the system, the initial recommendation of the system is obtained. The recommendation weight of the new user is the initial value, and the original recommendation is obtained; for the old user, the recommendation weight has been automatically updated according to the past browsing behavior. Therefore, the initial recommendation for logging in for the first time is to provide the recommendation service according to the user model.

The design of the system of this article is described in detail below. First, define the relevant variables as shown in Table 1.

| Variable   | Definition                                                                 |
|------------|---------------------------------------------------------------------------|
| User\(i\), \(i=(1,2,...,m)\) | User is user, \(m\) is number of users                                   |
| RS         | RS is Interest grade scale, Rating Scale                                 |
| \(r_j\), \(j=(1,2,...,n)\) | \(r\) is Interest degree, \(j\) is Interest level                      |
| \(X,x=1,2,...,x\) | \(X\) is Categories of information items, and \(x \in \mathbb{N}\)      |
| Item\(k_x\)=(Item\(1_x\),Item\(2_x\),...,Item\(k_x\)) | Item\(k_x\) is the \(k\) item                                           |
| R\(ik_x\)   | Represents the interest value of the \(i\)th User for the \(k\)th content |
| Time\(ik_x\) | Represents the browsing time of the first User for the \(k\)th item       |
| Hits\(ik_x\) | Represents the number of hits on the \(k\) item by the \(i\) User        |

The above variables are defined as the server-side access log, proxy log, and reference log in the web log form a recommended data source, and the user's preferences are determined below until a valid preference is implemented.

2.2 user similarity calculation

To record the users’ data, the user ID, corresponding browsing time, number of clicks and the user data is numbered by UserID, ItemNO, Time, and Hits, respectively. Among them, it is assumed that the user interest degree has five levels, that is, \(RS=(r_j)\), \(r_j=(1,2,...5)\), from which different interest degrees of different users can be distinguished. When preprocessing, the historical interest value must be converted to a level value that conforms to the format, that is, simple data preprocessing is performed. The user browsing behavior data table shown in Table 2:

| UserID | ItemNO | Time  | Hits   |
|--------|--------|-------|--------|
| User1  | Item1x | Time1x| Hits1x  |
| ...    | ...    | ...   | ...    |
| Usern  | Itemnx | Timeinx| Hitsinx|

This paper combines the two most basic features of the browsing behavior, Time, and Hits, which express the user session feature vector. Using the function formula (1), the degree of interest of the user \(i\) on the \(k\)th content can be judged.

\[
h(Times, Hits) = A \cdot Time_{i,k}^{x} + B \cdot Hits_{i,k}^{x} + C
\]  

(1)

In the formula, \(A\), \(B\), and \(C\) are a set of constants, and \(C\) is a correction coefficient (both empirical values).

The user's interest degree data table mainly records the browsing mode used, from which the
potential browsing habits can be found. As shown in Table 3:

| UserID | ItemNO | Rating |
|--------|--------|--------|
| User1  | Item1x | R1,kx  |
| ...    | ...    | ...    |
| Usern  | Itemkx | R1,kx  |

Standardize the user interest by Equation 2:

$$r_{\text{new}}^i = \frac{r_{\text{old}}^i}{RS_{\text{max}}^\text{old}} \cdot RS_{\text{max}}^\text{new}$$  \hspace{1cm} (2)

In the formula,
- $r_{\text{new}}^i$ indicates the updated user interest value;
- $r_{\text{old}}^i$ indicates historical interest value;
- $RS_{\text{max}}^\text{old}$ indicates the maximum value of the historical interest level scale;
- $RS_{\text{max}}^\text{new}$ indicates the maximum value of the updated interest level scale.

2.3 Implementation recommendations

The nearest neighbor of the target user is calculated by the above user preference similarity, and the next step is to generate a recommendation. For a target user $y_0$, formula 2 can be used to calculate his preference model, and then the similarity degree $w_{y_0, y}$ between $y_0$ and other users $y$ can be calculated. Finally, the preference possibility $P_{y_0}(x)$ of the target user $y_0$ for the unrated item $x$ can be expressed as:

$$P_{y_0} = \frac{\sum_{y \neq y_0} w_{y_0, y} \cdot (P_y(x) - \overline{P}_y)}{\sum_{y \neq y_0} w_{y_0, y}}$$  \hspace{1cm} (3)

If the final goal is to predict the target user's rating of the unrated item, the user's potential preference may be scaled accordingly, such as a number between the music network databases 1 to 5 to indicate the user's rating of the music. Then, the user's potential preference obtained according to the formula 2 can be multiplied by the corresponding number of scores to obtain a predicted score for the target user. When the final purpose is to recommend the target user, the user's preference for the item obtained according to formula 2 can be directly ranked in descending order, and the potential preference is top-N content as a result is recommended to the target user.

3. The Analysis of Experimental Results

3.1 Data selection

In order to verify the personalized recommendation based on user preference proposed in this paper, the Rizhao Vocational Technical College online on-demand movie scoring experiment data set is adopted, and the collaborative filtering recommendation algorithm based on user level is programmed [7]. The content-based collaborative filtering recommendation algorithm [8] and based on the user's potential preference algorithm, and then based on the commonly used evaluation criteria, the performance of the three algorithms is compared.

The data set is the Rizhao Vocational and Technical College online on-demand movie scoring database, which collects 73,021 ratings for 5,471 users for 2,546 movies. The user scores between 0 and 5 for an integer and a half, for a total of 10 levels. The higher the score, the higher the user's preference for the movie. At the same time, the database also stores 13,256 users accessing the site, including the time they spent visiting certain pages and the number of times they clicked on the movie.
3.2 Metrics [9]
The metrics for recommendation quality of e-commerce recommendation system mainly include statistical accuracy measurement methods (Statistical Accuracy Metrics) and decision support accuracy measurement methods (Decision Support Accuracy Metrics). The Mean absolute error (MAE) in the statistical accuracy measurement method is easy to understand, and can directly measure the recommended quality. It is the most commonly used recommendation quality measurement method. The average absolute deviation MAE measures the accuracy of the prediction by calculating the deviation between the predicted user score and the actual user score. The smaller the MAE, the higher the recommendation quality of the recommendation system.

In this paper, the average absolute deviation MAE is used as the metric. The data set is divided into two parts: training set and test set. The MAE is used to measure the items that the test set user has scored. Let the predicted test set a user score set be expressed as \( \{B'_1, B'_2, ..., B'_n\} \), corresponding to the actual user score set \( \{B_1, B_2, ..., B_n\} \), then the mean absolute deviation MAE is defined as:

\[
MAE = \frac{1}{n} \sum_{i=1}^{n} |R'_i - R_i|
\] (4)

Corresponding to the average absolute deviation MAE\(_i\) (\(i = 1, 2, ..., n\)) of all test set users, the average value can be calculated to obtain the total MAE value:

\[
MAE = \frac{1}{n} \sum_{i=1}^{n} MAE_i
\] (5)

In order to verify the recommendation superiority of the recommendation method based on the user's implicit preference, 200 user data were selected as the experimental object. The method based on user implicit preference is compared with other recommended methods. After the MAE calculation, the results are as shown in the test user MAE data sheet 4:

| Number of users | 10   | 50   | 100  | 150  | 200  |
|-----------------|------|------|------|------|------|
| User implicit preference recommendation | 0.2106 | 0.2023 | 0.1977 | 0.2042 | 0.2031 |
| Recommendation method based on user level | 0.2341 | 0.2284 | 0.2345 | 0.2325 | 0.2311 |
| Content-based recommendation method | 0.2265 | 0.2363 | 0.2290 | 0.2316 | 0.2384 |

As can be seen from the above data, the comparison of the same method decreases as the user's MAE value increases with the number of users, that is, the more the user's recommendation, the quality is higher. In different method comparisons, in the case where the number of users is the same, the MAE value of the recommendation method based on the user's implicit preference is significantly smaller than the other two methods, and this method is superior to the other two methods. In summary, the recommendation method based on the user's implicit preference has higher recommendation quality.

3.3 Obtained conclusion
For different data sets, in different cases, the average absolute deviation MAE of the method based on user potential preference similarity proposed in this paper is better than the collaborative filtering algorithm based on traditional similarity measure, and the recommendation quality is higher. This is because the method of this paper uses the potential preference of the user instead of the surface score, and makes better use of the existing information, and the calculation result is more practical and
accurate.

4. Conclusion
The algorithm uses the server-side log to analyze the length of the user's stay on the page it visits and the number of times a page is accessed instead of the explicit user score. Therefore, the amount of data obtained is much higher than the traditional collaborative filtering algorithm. According to statistics, browsing a product on the Internet is dozens of times more than the number of times the product is scored.

At the same time, traditional collaborative filtering recommendations are based on user-level ratings. Different rating scales of users have a greater impact on recommendations. Some users have strict ratings, some users have loose ratings, and similar surface scores may be two different types of users. In order to solve this problem, the algorithm uses a transformation model to convert the user's surface score into their potential preferences, so that the goals of the nearest neighbor of the target user can be found more accurately, thereby improving the credibility of the preference and improving the quality of the recommendation can be achieved.

REFERENCES:
[1] MA Li, TAN Wei, LI Pei. Research on user interest migration mode based on Web access information [J]. Computer Science. 2011(05)
[2] Timothy Rubin, America Chambers, Padhraic Smyth, Mark Steyvers. Statistical topic models for multi-label document classification [J]. Machine Learning. 2012(1)
[3] ZHONG Da-shuang. Research on knowledge service recommendation system based on Cloud Computing [D]. He'nan Normal University. 2013
[4] DUAN Jing-shan. Research on Web service composition method based on user preferences [D]. South-West University. 2014
[5] SONG Mei-qing. Research on multi granularity user preference mining for collaborative filtering recommendation [J]. Modern Library and Information Technology. 2015(12)
[6] LI Jian-ting, GUO Ye, TANG Zhi-jun. User interest computation based on user browsing behavior analysis [J]. Computer Engineering and Design. 2012(3)
[7] XING Chun-xiao, GAO Feng-rong, ZHAN Si-nan, ZHOU Li-zhu. Collaborative filtering recommendation algorithm based on user level [J]. Computer Research and Development, 2007(2)
[8] WANG Qiao, XIE Ying-hua, YU Shi-cai. Collaborative filtering algorithm combining user clustering and project type [J]. Application of Computer System. 2016(12)
[9] WANG Shi-dong. Design and implementation of e-commerce recommendation system based on Data Mining [D]. University of Electronic Science and technology of China. 2013