A Time-Related Composite Filtering Recommendation Method

Danlin Cai¹,²,³,⁴, Daxin Zhu¹,²,³,⁴, Junjie Liu¹,²,³

¹Quanzhou Normal University, China
²Fujian Provincial Key Laboratory of Data-Intensive Computing, China
³Fujian University Laboratory of Intelligent Computing and Information Processing, China
⁴Fujian Provincial Big Data Research Institute of Intelligent Manufacturing, China

Abstract—In this paper, a time-related composite filtering recommendation method is proposed, which can respond to the changing behavior of users and adjust the ranking of recommendation results in real time, so as to continuously improve the user’s experience in the recommendation system.

Keywords—Recommendation System, Time-Related, Collaborative Filtering

I. INTRODUCTION

With the popularity of the Internet and the rapid development of e-commerce, e-commerce system provides more and more choices, and its structure becomes more complicated, so users are often lost in a large amount of commodity information space, and unable to find the goods they need. In order to provide better service to users and meet the needs of users for personalized service, e-commerce recommendation system comes into being. E-commerce recommendation system solves the problem of “information confusion” and “information overload”, and realizes the direct interaction with users. It can simulate the store sales personnel to provide the goods recommendation to the users, and help the users quickly and accurately find the goods they really need. A number of survey results show that users are more willing to visit websites that provide personalized recommendation services.

Everything in the real world is constantly changing, and so is the recommendation system. User’s interest, and the attributes of the goods are constantly changing. Time is a kind of important context information, and the users have different interests in different time periods, for example, users’ interests in the daytime and at night may be different, their interests on weekdays and at weekends may be different, they may have a special interest in the festival and they may also have different interests in different seasons. Therefore, the rational use of time information and the consideration of the dynamic characteristics of recommendation systems will be of great help to improve the accuracy of recommendation and the satisfaction of users. The traditional recommendation system does not take the time factor into consideration when designing. Although users continue to have new behaviors and new items will be continuously added to the website, they do not take the initiative to consider the impact of time factors, so it can not give full play to the effect of time information on recommendation systems. With the increasing scale of e-commerce systems, recommendation systems face a series of challenges, especially the balance problems between the real time and recommendation quality. Recommendation system is difficult to meet the real-time requirements of the system when dealing with large-scale data. Most recommendation techniques ensure the real-time requirements, and meanwhile, take the sacrifice of the recommendation quality of recommendation system as the premise. In a large scale of reality, if the user’s recommendation needs can not be met in time, it will undoubtedly affect the user’s evaluation of the system. Therefore, it is necessary to do further research on how to effectively improve the recommendation quality of recommendation systems while ensuring real-time recommendation services. The research
of real-time dynamic recommendation systems with high accuracy has become a new subject of the current recommendation system research.

At present, the development of e-commerce in China is changing the traditional trade mode at a great speed, and the development of recommendation system is also emerging. Many large e-commerce websites are successful because of the implementation of recommendation systems, such as Taobao, Jingdong Mall, Dangdang and so on. Through the recommendation system, enterprises not only provide users with a humanized service experience, obtain the user’s favorable impression, but also bring a win-win profit. However, in the implementation and application of e-commerce recommendation systems, compared with foreign e-commerce websites, the gap between domestic e-commerce websites and foreign e-commerce websites is still relatively large, which is mainly reflected in the following aspects:

The first point is the lack of personalized recommendation. Many e-commerce websites usually recommend all hot commodities to users, and allow users to choose the commodities they want to buy, so the effect can be imagined.

The second point is that the dynamic recommendation degree of the timeliness is relatively low. Many e-commerce website recommendation systems can not retain the user’s information very well, and use the user’s information. Overall, the recommendation strategy basically stays at the level of finding.

The third point is the single recommendation strategy. There is a lack of hybrid recommendation technology. In particular, there is a lack of bidirectional complementarity between content-based recommendation technology and collaborative recommendation technology.

II. TRADITIONAL RECOMMENDATION TECHNOLOGY
2.1 Content-based recommendation technology

Content-based recommendation technology is relatively simple and effective in text content recommendation, and the recommendation results are more intuitive and need no domain knowledge. The basic idea is to recommend a given user with other items that are similar to the content of a item that he liked before[1]. This recommendation only needs two kinds of information: the description of item characteristics and the user’s past preference information. It does not require a large user community or scoring history, and it can produce a list of recommendations for a single user. The key technologies are modeling the item and modeling the preferences of users, and calculating their similarity. There are many kinds of modeling methods. The vector space model proposed by Salton and others is the most commonly used recommendation method based on content data[2]. There are many other modeling methods, such as improved vector space model, explicit decision making model, linear classification, machine learning and so on.

There are two major flaws in content-based recommendation: The first point is that the item recommended to the user is very similar to the one that the user has already consumed, which makes it difficult to discover the types of items that the users are not familiar with but potentially interested in, and the surprise is not high. The second point is the need to preprocess the item to get the item characteristics that can represent them, but this pretreatment is often very difficult in practical problems, especially the data in the field of multimedia(image, audio, video, etc.). In view of the above shortcomings, this paper puts forward the method that uses domain experts to label the item[3], that is the traditional classification system. This method is more scientific and authoritative, but it can not express the user’s personal views well, so it is one-way without interaction. The other is that users label the item, that is the mass classification system. It is more diversified, not accurate, but can timely feed back users’ views and help to dig interest[4].


2.2 Recommendation technology based on collaborative filtering

Collaborative filtering technology is one of the most widely used and successful recommendation technologies in recommendation systems. It does not require the description of commodity characteristics, it learns about the similarity between user purchase behaviors, and it is not dependent on the characteristics of commodities, so it can recommend the commodities that look different from the surface features, but have great relevance in fact.

Collaborative recommendation algorithm directly uses the historical scoring of users or items, calculates the similarity of users or items according to the scoring, and selects the users or items with the most similarity as the nearest neighbor, and then predicts the possible scoring of users according to the scoring of the nearest neighbor, so as to generate the recommendation. It can be divided into three steps:

1. The first one is the data representation. Through the collected scoring information of users for items, a m×n matrix can be constructed to represent, as shown in the table listed above. There are m users and n items in the matrix, and the value in the i row and j column represents the scoring of the item j by the user i. If the users prefer the item, the corresponding score is high, and if they don’t like it, the score is low. The score generally ranges from 1 to 5.

   The second one is to calculate the correlation similarity between users. In order to find the nearest neighbors, we need to measure the similarity between users. The higher the similarity, the more similar the personality between users and the preferences of users for commodities. The similarity between users is usually calculated by Pearson correlation coefficient and cosine similarity. Assuming that the current user is i and the target user is j, the formula of Pearson correlation coefficient sim (i, j) of user i and j is formula (1).

   \[
   \text{sim}(i, j) = \frac{\sum_{c \in I_{i,j}} (R_{i,c} - \bar{R}_i)(R_{j,c} - \bar{R}_j)}{\sqrt{\sum_{c \in I_{i,j}} (R_{i,c} - \bar{R}_i)^2} \cdot \sqrt{\sum_{c \in I_{j,c}} (R_{j,c} - \bar{R}_j)^2}}
   \]

   In the formula, \( R_{i,c} \) and \( R_{j,c} \) respectively represent the scoring value of the item c by the user i and user j, \( \bar{R}_i \) and \( \bar{R}_j \) respectively represent the average scoring of user i and user j in data scoring matrix, and \( I_{i,j} \) represents the item set scored by the user i and user j together.

   The third one is to calculate the target users’ prediction scoring for the item. k nearest neighbors should be found, and meanwhile, the corresponding k similarity \{sim1, sim2,..., simk\}can be obtained. Through the scoring of the item by k nearest neighbors and the weighting of the similarity between the target user and the nearest neighbor, the prediction scoring of the item by the target user is obtained, and the calculation method is formula (2).

   \[
   P_{u,i} = \bar{R}_u + \frac{\sum_{n\in NN_u} \text{sim}(u, n) \times (R_{n,j} - \bar{R}_n)}{\sum_{n\in NN_u} \text{sim}(u, n)}
   \]

   In the formula, \( \text{sim}(u, n) \) represents the similarity value between user u and user n, \( NN_u \) represents the nearest neighbor user set of the target user u, \( R_{n,j} \) represents the scoring of the item i by the nearest neighbor n, and \( \bar{R}_u \) represents the average value of the item of all neighbor users.
III. TIME-RELATED COMPOSITE FILTERING RECOMMENDATION METHOD

We propose a time-related composite filtering recommendation method that incorporates the features of content-based recommendation techniques into collaborative filtering, which not only overcomes the sparsity problem in collaborative filtering but also eliminates the need for co-filtering which is only recommended when it is rated as high score by the nearest neighbors. The time-related composite filtering recommendation method can be directly recommended when the item matches the user profile information. At the same time, in order to make accurate recommendations in different time-sensitive websites, we use a time-graph model that models both social and personal factors by introducing two new nodes and adjusting the weight of the two nodes to adjust the accuracy of real-time recommendation system. The main contents include:

The first one is the behavior extraction and analysis.

The user behavior is extracted from the user behavior database, and the user behavior is analyzed to obtain the user’s interest characteristics. Users behave in a variety of ways on a website, each of which contains the user’s interest in items. At the same time, according to the theory of time effect, the time when user behavior occurs also has a very important influence on the final recommendation result, and the recent behavior of the user more reflects the user’s recent interest. The user \( u \) is given, and we let \( N_b(u) \) be the item that user \( u \) touches in behavior \( b \), and items set \( N(u) \) that the user \( u \) is interested in is defined as formula (3).

\[
N_b(u) = \bigcup_b N_b(u)
\]  

We introduce the time attenuation coefficient to extract and analyze the behavior, and use formula (4) to measure the impact on the behavior after the introduction of time.

\[
N_b(u) = \sum_{b,i\in N_b(u)} \lambda be^{-2\beta(t_b-t_s(u,i))}
\]  

We experimentally study the difference between the attenuation coefficients of different product categories and propose a multi-dimensional behavior extraction method.

The second one is the real-time dynamic recommendation engine.

The engine is responsible for contacting the user’s interest vectors and items to be recommended. If the user’s interest characteristics only include the user’s interest in certain items, then the relevant recommendation module mainly depends on the correlation matrix between items. The main recommendation of the relevant recommendation process is to search the correlation matrix for each type of items that the user likes, so as to find the most similar \( M \) items, and finally all the items will be aggregated by relevancy to become the recommended result. The recommended engine modules in the system often rely on the related tables. Each related table defines the correlation between the items. We let \( W_k \) be the \( k \)-th correlation table, \( w_k(i,j) \) defines the item \( i,j \) is the similarity on the \( k \)-th correlation table, \( S_k(i; M) \) is a collection of \( M \) items that are most similar to the object \( i \) in the \( k \)-th correlation table, and the user’s preference is calculated by the formula (5).

\[
p(u,i) = \sum_k \sum_{j\in N(u)} w_k(j,i)r(u,j)
\]
After extracting the multi-dimensional behavior, the accuracy of the real-time recommendation system is dynamically adjusted through the time map model. The real-time dynamic recommendation effect is judged by the accuracy, coverage and diversity.

IV. EXPERIMENTAL VERIFICATION

4.1 Evaluation criteria
Recall rate (RECALL), precision (PRECISION), mean average error (MAE), and average absolute user error (MADE), these are the recommended systems to measure the quality of the recommended quality measures. These methods are more intuitive and easy to understand. In this paper, we use the recall rate (RECALL), precision (PRECISION), and mean average error (MAE) to measure the accuracy of the forecast.

Formula (6) is the recalculation method of RECALL.

\[
RECALL = \frac{\sum_{u \in U} |R(u) \cap T(u)|}{\sum_{u \in U} |T(u)|}
\] (6)

Formula (7) is the calculation method of PRECISION.

\[
PRECISION = \frac{\sum_{u \in U} |R(u) \cap T(u)|}{\sum_{u \in U} |R(u)|}
\] (7)

Here R (u) denotes the recommended list of user u, T (u) denotes the list of behavior records of user u in the test set, and the larger the values of these two indexes are, the better the algorithm is.

Equation (8) is the calculation method of MAE

\[
MAE = \frac{\sum_{i=1}^{N} |p_i - q_i|}{N}
\] (8)

Here, \(p_i\) is the predicted score set, \(N\) is the predicted number of items, \(q_i\) is the actual score set, which reflects the recommended accuracy by calculating the deviation between the predicted and actual values. The smaller the MAE value is, the closer the score is to the true score, the more accurate the forecast is and the better the final recommendation is.

4.2 Experimental process and conclusions
This paper uses the data set provided by the MovieLens site for a research-based recommendation system based on Web. This site collects registered users’ rating data on movies and user-recommended movie listings. It includes 706 users’ evaluations for the movie, and each user evaluates at least 20 movies. The evaluated values range from 1 to 5. If users are more satisfied with the movie, then the score is higher. The data is preprocessed to remove noise and other disturbances and turn it into clean data. The dataset is divided into three datasheets: user attributes, movie attributes, and user rating scores for the movie. Among them, the user attribute table includes user number, gender, age, occupation and other attributes. The movie property sheet includes information such as movie number, movie title, and movie genre. The user’s scoring data table for the movie includes a user number, a movie number, a user’s rating on the movie, and time stamp information. We compare the content-based recommendation method BMR, the collaborative filtering-based recommendation method BFR, and the recall(RECALL), precision(PRECISION) and mean absolute
error (MAE) of the time-related composite filtering recommendation TBRR proposed by us in the MovieLens test.

Table 1. Comparison of BMR, BFR and TBRR test indicators in the MovieLens test set

<table>
<thead>
<tr>
<th></th>
<th>BMR</th>
<th>BFR</th>
<th>TBRR</th>
</tr>
</thead>
<tbody>
<tr>
<td>RECALL</td>
<td>0.873</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>PRECISION</td>
<td>0.725</td>
<td>0.781</td>
<td>0.790</td>
</tr>
<tr>
<td>BMR</td>
<td>0.865</td>
<td>0.852</td>
<td>0.832</td>
</tr>
</tbody>
</table>

It can be seen that the TBRR proposed by us combines the advantages of content recommendation technology and collaborative filtering technology, and introduces the time parameters to improve the real-time recommendation, and the related technical indicators are superior to the traditional BMR and BFR technologies.

V. CONCLUSION

E-commerce plays a very important role in promoting “the second undertaking” of private enterprises, changing the mode of economic development, and promoting the sustainable development of the economic society in Quanzhou. And the application of efficient real-time dynamic recommendation system in e-commerce will be an important way to meet and enhance consumer demand, improve the optimization degree of the industry and resources, and change the way of economic development. Compared with the traditional content-based recommendation method and the recommendation method based on the collaborative filtering, the time-related composite filtering recommendation method proposed by us has the following two advantages: The first one is to dynamically adjust the accuracy of the real-time recommendation system by using the time graph model in combination with the real-time problem of the recommendation system. The second one is to respond to the changing behavior of users and adjust the ranking of recommendation results in real time, so as to continuously improve the user’s experience in the recommendation system. The experiments results show that this method can respond to the changing behavior of users and adjust the ranking of recommendation results in real time, so as to continuously improve the user’s experience in the recommendation system.

ACKNOWLEDGEMENTS

The study is supported by Natural Science Foundation of Fujian Provincial Science and Technology Department (No.2017J01508) and Key Projects of Fujian Provincial Education Department (No.JA15388) and Quanzhou Science and Technology Project (No.2014Z133) (NO.2012Z101).

REFERENCES