Research on collaborative filtering based on user interest in the higher vocational e-commerce website development

Songjie Gong*
Zhejiang Business Technology Institute, Ningbo, 315012, China

*Corresponding author: songjie_gong@163.com

Abstract. E-commerce Recommender Systems suggest useful and interesting products to customers in order to increase user satisfaction and online conversion rates. They typically exploit explicit or implicit user feedback such as ratings, buying records or clickstream data and apply statistical methods to derive recommendations. Collaborative filtering is one of the most widely used approaches in recommender systems. However, the traditional collaborative filtering methods compute users' similarities in the dimension of products, and they do not take the influence of neighbor users into consideration when computing such similarities. This paper focuses on explicitly formulated user requirements as the sole type of user feedback. The new collaborative filtering method based on user interests’ transmission. This method computes customers’ similarities in the dimension of interests, and takes the interests transmission between different customers into reflection. This method not only can cope with cold start problem and data sparse problem, but also have prediction precision.

Keywords: collaborative filtering, user interest, e-commerce

1. Introduction
According to the interaction history between users and the system and the user's personal information, the recommender system constructs the user's interest model and predicts the products that users may be interested in. It not only saves the user's time and improves the probability of product cross selling, but also enhances the user's loyalty and purchase desire by providing products that meet the user's requirements. At present, there are many meaningful researches on recommendation system, including content-based recommendation, collaborative filtering based recommendation and hybrid recommendation.

Among them, the method based on collaborative filtering only makes recommendations according to the interaction history between users and the system and the preferences of other users, and does not need to mine the content information of products and users (for example, keywords representing products or personal information of users, etc.). Compared with the other two methods, CF has less restrictions and better recommendation effect, so it is widely used Amazon.com Using CF to recommend books, CDs and other products, GroupLens recommends news and movies for users.
CF can also be divided into user based collaborative filtering and item based [13] collaborative filtering. The former thinks that a given user will like the products that users with similar preferences like, while the latter thinks that a given product will be liked by users who like the products similar to them.

2. Traditional personalized recommendation system

Collaborative filtering is one of the earliest and most successful technologies in personalized recommendation system. Collaborative filtering, also known as social filtering, its basic idea is very intuitive: in daily life, people often make some choices (shopping, reading, music, etc.) according to the recommendations of their relatives and friends. The traditional collaborative filtering recommendation technology generates recommendation results according to the user's explicit score. The starting point of collaborative filtering technology is that everyone's interest is not unique and unpredictable, and the individual's preference is often within a certain group. Therefore, the key to evaluate the corresponding resources according to the users with the same or similar interests and recommend them to other users is the discovery of similar user groups. At present, the research of collaborative filtering technology mainly includes user based collaborative filtering technology and Project-based Collaborative filtering technology.

Problems in traditional collaborative filtering recommendation technology

(1) User data is obtained by explicit scoring. The obvious disadvantage of explicit rating is that users have to pause browsing or reading and enter the rating instead, which is not in line with the habit of most users. Unless we know clearly the benefits of scoring web pages (such as getting small gifts, etc.), most users are not willing to waste time doing such meaningless work, which will lead to the lack of scoring data. Experiments show that only when each web page has a considerable amount of rating data, the recommendation system can produce more accurate recommendation results, and the extreme sparsity of user rating data will directly lead to the decline of recommendation quality.

(2) We can't find new interesting resources for users, we can only find resources similar to users' existing interests.

(3) There is no comprehensive user personalized search and the use of group commonness to do active recommendation.

(4) Poor scalability, and with the further expansion of the scale of the system, the number of users and project data increase sharply, which will also lead to extreme sparsity of user rating data. In the case of extremely sparse user rating data, the traditional similarity measurement methods have their own disadvantages, which makes the nearest neighbor of the target user calculated inaccurate, and the recommendation quality of the recommendation system drops sharply.

3. Personalized recommendation model based on user interest

In view of the problems existing in traditional collaborative filtering technology, a personalized recommended UPR (users' interest based personalization recommendation) model based on user interest is proposed. UPR model uses the user interest contained in the network log to design a user interest model based on the frequency of web users' access. The assumption of the model is that a group of pages visited by people with similar interests may be related. The model finds out the users interested in the input page from the user group, and finds out a kind of users with similar interest background and the most interested in the input page through clustering. The paper combines the pages of interest of such users, and then excavates the pages related to the input page. The analysis object of the model is the user access frequency matrix, which is called user interest matrix in the model. The user's interest in the page is defined based on the user's access frequency to the page. The recommended system model is shown in Figure 1.
Figure 1 personalized recommendation model based on user interest

Compared with traditional collaborative filtering technology based on neighbor users, the personalized recommendation model based on user interest has the following improvements:

(1) Most information filtering systems need users to participate in the evaluation of objects, and use the evaluation of these displays to predict. The information used by UPR model is network log, which is left by users unintentionally. Therefore, the model does not need the participation of users, all the steps are completed automatically by the system, which overcomes the problem of data sparsity in the initial stage of information filtering system.

(2) Another fundamental difference between UPR model and CF model is the application of clustering in specific algorithm. CF model only does user interest clustering or item clustering, and UPR model can be regarded as the interest clustering for users first, then the web page is related clustering based on the results of interest clustering.

(3) In the specific steps of clustering users' interest, CF model and UPR model also have differences. The user clustering in CF technology is to find a kind of users with similar interests centered on new users. UPR model is to cluster interests among a group of users. The center of clustering and the result of clustering are unknown before clustering. Therefore, it is more difficult to cluster UPR model.

(4) CF model is the evaluation of a group of objects P by a group of users with similar interest. For a new user UK, if it is judged that the new user UK is similar to the user group u, the system can predict the evaluation of UK on P based on u-evaluation of P. UPR model is known a group of users with similar interests u to evaluate a group of objects P, and a new user UK and user group u evaluate a new object pl. then the model can mine the objects related to the new object pl in that group of objects P, and push the relevant pages to users who visit pi. Compared with CF model, UPR model recommends more accurate objects.

4. Collaborative filtering based on user interest algorithm

In order to effectively prevent the problems of traditional collaborative filtering such as high product dimension and data sparsity, this paper proposes a collaborative filtering algorithm uit based on user interest propagation. In this section, the uit algorithm is introduced in detail. The recommendation algorithm based on graph structure points out that bipartite graph can be used to establish user product relationship.

User set u and product set P are defined, and user product bipartite graph with set $u \cup P$ as vertex set is constructed. Among them, if the user u hits the product P excessively, and the score value is $R_{UP}$, an edge is connected between u and P, and the edge weight $A_{UP} = R_{UP}$ ($s \in u, j \in P$). The bipartite graph is projected on the user dimension to get the corresponding user projection graph, namely user association graph. In the user projection graph, the weight $W(s, j)$ of user u to V represents the influence degree of user u on user V preference.
\[ w(s, j) = \frac{1}{k_s} \sum_{p=1}^{k_s} a_{sp} a_{jp} \]  

Among them  
\[ k_s = \sum_{p=1}^{k_s} a_{sp} \cdot k_p = \sum_{p=1}^{k_s} a_{jp} \]

Define the user incidence matrix RM \( |s| \times |j| \) (relation matrix), let \( rmuv = w(s, j) \), and normalize \( rm \) with behavior unit, then RM corresponds to the incidence matrix of user incidence graph. Among them, \( rmuv \) is the correlation coefficient of user \( u \) to user \( V \), that is, the proportion of user \( U \)'s influence among all users' influences on \( V \) preference.

A \( k \)-dimension interest vector is defined for each user, and each user is a probability distribution on the interest vector. The interest vector is normalized, that is, for any user, the sum of the probabilities that he belongs to each interest is 1. In uit, each user is randomly assigned an initial interest distribution. The matrix \( im|x|K \) can be used to save the \( k \)-dimensional interest vectors of all users, which is called user interest matrix. The initial random interest distribution obviously can not reflect the real interest of users. Therefore, we simulate the real scene and think that the user's interest is affected by the friends around us. We update the user's interest distribution by letting the user's interest vector randomly walk on the user association graph to learn the user's real interest. Suppose that after step I random walk, the user interest matrix is updated to \( IM^0=RM \bullet IM^{i-1} \)

Where, \( IM^0 = IM \). After step random walk, the final user interest matrix is \( im \), namely \( IM^{step} \).

In the user based CF algorithm, it is necessary to find the neighbor set similar to the given user by calculating the similarity between users. Traditional coordination filtering methods often use Pearson correlation or angle cosine correlation based on product dimension to calculate user similarity [7]. However, the product vector is often very high and sparse, and only the direct similarity between two user vectors can be considered when calculating the user similarity, not the indirect influence of other users. Therefore, there is a big deviation between the similarity results calculated by this method and the actual similarity of users.

In order to describe the similarity between users more accurately, uit uses Pearson correlation based on interest dimension or cosine correlation based on interest dimension to calculate the similarity between users.

The Pearson correlation between user \( u \) and \( V \) based on interest dimension is defined as:
\[
sim(u, v) = \frac{\sum_{i=1}^{G} (IM_{u,i} - \bar{IM}_u)(IM_{v,i} - \bar{IM}_v)}{\sqrt{\sum_{i=1}^{G} (IM_{u,i} - \bar{IM}_u)^2} \sqrt{\sum_{i=1}^{G} (IM_{v,i} - \bar{IM}_v)^2}}
\]

5. Summary

Although CF method is simple and practical, the traditional user based CF method has been limited by two aspects: first, the user similarity needs to be calculated in the product dimension, and the vector composed of the product is often very high and sparse, so there is a certain deviation between the calculated similarity result and the actual similarity of the user; second, when calculating the user similarity, only the direct similarity of two user vectors can be considered Consider the indirect impact of users. In real life, a user's interest is easily affected by the friends around him, if the calculation process cannot take these indirect factors into account. This paper focuses on explicitly formulated user requirements as the sole type of user feedback. The new collaborative filtering method based on user interests’ transmission. This method computes customers’ similarities in the dimension of interests, and takes the interests transmission between different customers into reflection. This method not only can cope with cold start problem and data sparse problem, but also have prediction precision.

References
[1] Adomavicus G, Tuzhilin A. Toward the next generation of recommender systems: A survey of the
state-of-the-art and possible extensions [J]. IEEE Trans on Knowledge and Data Engineering, 2005, 17(6):734-749.

[2] Songjie Gong. A Collaborative Filtering Recommendation Algorithm Based on User Clustering and Item Clustering, Journal of Software, Volume 5, Number 7, July 2010, pp: 745-752.

[3] Linden G, Smith B, York J. Amazon.com Recommendations: Item-to-item collaborative filtering [J]. IEEE Internet Computing, 2003(1):76-80.

[4] Resnick P, Iacovou N, Suchak M, et al. Grouplens: an open architecture for collaborative filtering of Netnews [C]// Proceedings of the CSCW conference, Chapel Hill, NC, 1994:175-186.

[5] Debnath S, Ganguly N, Mitra P. Feature weighting in content based recommendation system using social network analysis [C]// Proceedings of WWW, Beijing, China: April, 2008:1041-1042.

[6] Balabanovic M, Shoham Y. Fab: Content-based, collaborative Recommendation [J]. Comm. ACM, 1997, 40(3):66-72.