Improving Recommender System's Efficiency Using Time Context and Group Preferences

Mohammad Julashokri, Mohammad Fathian, Mohammad Reza Gholamian, Ahmad Mehrbod

Advances in Information Sciences and Service Sciences. Volume 3, Number 4, May 2011

Improving Recommender System's Efficiency Using Time Context and Group Preferences

Mohammad Julashokri,
Mohammad Fathian,
Mohammad Reza Gholamian,
Ahmad Mehrbod

1, First Author
Iran University of science and technology, julashokri@ind.iust.ac.ir
2, Iran University of science and technology, fathian@iust.ac.ir
3, Iran University of science and technology, gholamian@iust.ac.ir
4, Isfahan University of technology, mehrbod@cc.iut.ac.ir
doi:10.4156/aiss.vol3.issue4.20

Abstract

By the expanse of internet stores and products, recommender systems have emerged to increase store attractiveness and develop online customers. Recommender systems are systems which help customers to find product that they want. These systems recommend product to individual customer according to their preferences and interests. Recommender systems use several ways such as collaborative filtering and content-based filtering to create recommendation. In this study we proposed a recommender system based on collaborative filtering. In proposed model we endeavored to improve the customer profile in collaborative systems to enhance the recommender system efficiency. We do this improvement using time context and group preferences.

Keywords: Recommender System, Collaborative Filtering, Customer Profile

1. Introduction

Since old days, there were different kinds of marketing including mass marketing, segmentation marketing, one to one marketing, etc. One to one marketing is an effort to find customer individual needs and respond to them. This individualization can be in products or in the way of offering products. Recommender systems have emerged in e-commerce applications to support product recommendation, which provide one to one marketing (Schafer et al., 2001; Liu & Shih, 2005). Indeed, recommender systems individualized the way of recommending products. These systems try to recommend different products to each customer with collecting data of customer preferences and data mining techniques.

The recent commercial success of recommender systems has been appeared in many online stores such as Amazon.com, CDNow.com, Barnes&Noble.com, and MovieFinder.com. Based on how recommendations are generated, recommender systems are usually classified in to content-based filtering (CBF) (Cheung, Kwok, & Law, 2003; Weng & Liu, 2004; Cho & Kim, 2004; Hung, 2005; Adomavicius & Tuzhilin, 2005; Leung, Chan, & Chung, 2006; Shih & Liu, 2008), collaborative filtering (CF) (Karypis, 2001; Sarwar et al., 2001; Cheung et al., 2003; Weng & Liu, 2004; Cho & Kim, 2004; Liu et al., 2005; Adomavicius & Tuzhilin, 2005; Boucher-Ryan & Bridge, 2006; Leung et al., 2006; Shih & Liu, 2008) and hybrid approaches (Burke, 2002; Cho & Kim, 2004; Semeraro et al., 2005; Adomavicius & Tuzhilin, 2005; Choi et al., 2006; Kim, Li, Park, Kim, & Kim, 2006; Shih & Liu, 2008; Albadvi and Shahbazi, 2009).

The rest of this study is organized as follows. Section 2 outlines the background and reviews related work on customer profile and recommendation methods. Section 3 illustrates the proposed methods. Finally, Section 4 draws conclusions and summarizes the contributions of this work.

2. Research background

In this section we review previous recommendation methods and customer profile.
2.1. Customer lifetime value and RFM evaluation

Customer lifetime value is typically used to identify the customers' value for a business and show profitable customers. Businesses can develop marketing strategies for target customer with measuring customer lifetime value because different kind of marketing must be executed for customers with different customer lifetime value. Some studies like Liu and Shih (2005) employed customer life time value for product recommendation. Measuring RFM variables is a proper way to assess customer lifetime value (Liu & Shih, 2005). Bult and Wansbeek (1995) described the RFM terms as follows: (1) \( R \) (Recency): time period since last purchase occurred. A lower value of recency means a higher probability of purchase repetition; (2) \( F \) (Frequency): number of purchases occurred during defined period. A higher value of frequency means a higher loyalty; (3) \( M \) (Monetary): the amount of money spent by customer during the last defined period. A higher value of monetary means a higher profitability.

This metrics usually were not ready at first and must be extracted from raw data for data mining process. Some studies employed RFM metrics in recommendation process. Liu and Shih (2005) proposed an approach that clustered the customer with RFM metrics and mined association rules in each cluster for recommendation generation.

2.2. Recommendation methods

In the literature, two types of recommendation methods, content-based filtering and collaborative filtering, and a hybrid of them have been more suggested (Resnick & Varian, 1997; Wang, Chuang, Hsu, & Keh, 2004). Content-based filtering is based on a comparison between content of items and a user profile (Wang & Shao, 2004). In content-based filtering method characteristics of user interesting items are extracted and other items with these characteristics are recommended. Systems analyze the content of items and create customer profiles that are a representation of a user’s interest in terms of keywords, phrases, and features. Then, the systems with a comparison between user profile and content of items unknown to the user, estimate which of items could be interesting to the user (Min & Han, 2005; Park and Chang, 2009).

Collaborative filtering has been more used in ecommerce applications. Content-based filtering computes similarities between data items and user profiles, but collaborative filtering considers similarities between user profiles. In collaborative filtering a user’s profile consists of user specified ratings to items. These ratings are compared to other users to find similarities between users. Ratings of new items are predicted using a combination of the ratings of the nearest neighbors to those new items. Nearest neighbors of a user are defined using the correlation between past ratings (Montaner, Lopez, & De La Rosa, 2003; Park and Chang, 2009). We would more explain item-based collaborative filtering because our proposed model is based on this method.

2.2.1. Item-based collaborative filtering

Collaborative filtering is powerful recommendation method which has been widely used in retail marketing applications. An item-based collaborative filtering recommends products to target customer based on similarities between items (Sarwar et al., 2001). Item-based collaborative filtering finds products similar to products purchased by target customer then computes the interest level of target customer on those products and finally recommends the top \( N \) (\( N \) most similar) products to target user.

The item-based collaborative filtering algorithm is detailed as follows. At first, customer profile matrix (R) must be prepared. This matrix shows the customer preferences or customer purchasing history. \( r_{ij} \) represents the interest level of \( i \)th customer in \( j \)th item. In different cases elements of profile matrix are different; in some case such as study of Sarwar et al.(2000), matrix elements were numbers between 1 to 5 which \( r_{ij} \) show the interest level of \( i \)th user in \( j \)th movie and, zero if \( i \)th user has not rated \( j \)th movie. In Liu and Shih (2005) study elements were binary that \( r_{ij} \) is one if the \( i \)th customer purchased the \( j \)th product, and otherwise is zero. Next step in item-based collaborative filtering is to measure similarity between items that can be computed in several ways. One common method to
compute similarity between items \( i \) and \( j \) is adjusted cosine which is defined as follows (Sarwar et al., 2001):

\[
\text{sim} (i, j) = \frac{\sum_{u \in U}(R_{u,i} - \bar{R}_u)(R_{u,j} - \bar{R}_u)}{\sqrt{\sum_{u \in U}(R_{u,i} - \bar{R}_u)^2} \sqrt{\sum_{u \in U}(R_{u,j} - \bar{R}_u)^2}}
\]

(1)

Here, \( R_{u,i} \) and \( R_{u,j} \) are elements of profile matrix showing the interest level of \( u \)th customer in \( i \)th and \( j \)th products respectively. \( \bar{R}_u \) is the average of the \( u \)th user's ratings. Only users who have rated both items \( i \) and \( j \) are participated in similarity computation (figure 1).

![Figure 1. Isolation of the co-rated items and similarity computation (Sarwar et al., 2001)](image)

Next step is prediction of interest level of target customer in products which were obtained from last step. This method predicts the interest level of \( u \)th user in \( i \)th item by computing the sum of the ratings given by user on the items similar to \( i \). Each ratings should be multiplied to similarity between items \( i \) and \( j \). The method predicts interest level of user \( u \) in product \( i \) (\( P_{u,i} \)) as follows (Sarwar et al., 2001).

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

(2)

Here, \( \text{sim}(i,j) \) is similarity between items \( i \) and \( N \) (\( N \) means each item which is similar to \( i \)) and \( R_{u,N} \) is interest level of user \( u \) in item \( N \).

**2.3. Customer profile**

Usually, in collaborative filtering method, similarities between user profiles are computed for recommendation generation. Thus, a proper profile increases the precision of recommender system. Lee and Yang (2003) suggested that important factor in personalization of researches is created the computational model for each user to predict his preferences. Park and Chang (2009) noted that "The development of efficient customer profile models is crucial for improving the recommendation quality of the recommendation system ".

In most recommender systems user profile is a matrix that explicitly indicates the users' interest or preference in items (products). Albadvi and Shahbazi (2009) created the user profile structure in their recommender system called HRPCA that is matrix of matrixes. Their user profile contained users ratings on attributes of each grain category. Elements of user profile matrix represent the level of interest of a user in an item. User ratings can be appointed directly from customer or constructed from shopping steps (Cho & Kim, 2004; Albadvi & Shahbazi, 2009; Park & Chang, 2009):

\[
r_{\text{q}} = \frac{r_{\text{b}}}{r_{\text{t}}} + \frac{r_{\text{p}}}{r_{\text{t}}} + \frac{r_{\text{p}}}{r_{\text{t}}}
\]

(3)
$r^c_i$ is the total number of clicks of user $i$ across product $j$ and $r^c_i$ is the total number of clicks of a user $i$. $\frac{r^p_i}{r^c_i}$ and $\frac{r^b_i}{r^c_i}$ are respectively defined as the total number of occurrences of basket placements and purchases of user $i$ across product $j$.

In the cases that ratings are directly appointed by user, usually, the precision is low and numbers must be improved. There are several methods to normalize the ratings. Park and Chang (2009) suggested that group behavior of customers is as important as their individual behavior.

3. Proposed model

In this study we propose a five-phase recommendation technique including data preparation, clustering, profile creation, profile improvement, and recommendation generation. In our proposed model for recommending the products, a hybrid approach based on collaborative filtering is used. Figure 2 shows the proposed approach generally.

3.1. Phase 1: data preparation

In the first step data would be prepared. There are vast amount of data in real world that must be cleaned, reorganized, reformatted, etc for using in data mining applications. In this phase, data must be transformed in new format appropriate for application. RFM variables must be extracted in this phase.

3.2. Phase 2: clustering

In this phase data would be clustered with SOM algorithm. Input fields of algorithm are age, gender, and occupation. SOM clustering would determine the number of clusters; then with that specified number k-means clustering would be executed. The cluster number of each customer would be used as her/his class label.

3.3. Phase 3: Profile creation

In this step user profiles would be created. In collaborative filtering user profile usually is matrix that its rows show users and its columns show items.
3.4. Phase 4: Profile improvement

The novelty of our work lies in this phase. In literature, there are several ways to improve customer profile. Park and Chang (2009) proposed that in addition to individual behavior, group behavior must be participated in customer profile. In our proposed model groups of users are clusters that were created in phase 2. Group preferences would be used to improving the customer profile at this way:

\[ t_r^i = \alpha * r^i + \beta * \bar{r}_{Gij} \] (4)

Where, \( r^i \) represents the element of profile matrix that shows the interest level of customer \( i \) in product \( j \) and \( t_r^i \) represents the element of improved profile matrix. \( \bar{r}_{Gij} \) is average of interest levels of group of customer \( i \) in product \( j \). Weighting values \( \alpha \) and \( \beta \) represent relative importance between individual and group preferences. For new customers \( \alpha = 0 \) and \( \beta = 1 \); but for other customers these coefficients must be appointed.

In our proposed model, according to this theory that whatever customer life time value is higher its preferences are more important, RFM variables are exploited to determine the \( \alpha \) and \( \beta \):

\[ \alpha + \beta = 1 \] (5)

\[ \alpha = 0.33 * \frac{30 - R}{30} + 0.33 * \frac{F}{m \cdot \mu(F, \bar{F})} + 0.33 * \frac{M}{m \cdot \mu(M, \bar{M})} \] (6)

\( R \) is number of days since last purchase. If \( R \) is greater than or equal to 30 the fraction would be changed to this:

\[ \frac{60 - R}{60} \] (7)

At this way if \( R \) is greater than or equal to 60, 90, and ... the fraction would be changed accordingly. \( F \) is number of products purchased by customer in the last month and \( \bar{F} \) is average of \( F \)s. Finally, \( M \) is number of products purchased by user in last session and \( \bar{M} \) is average of \( Ms \).

The other way to improve the customer profile is using context. According to theory that customer preferences may be changed during time, we used time context to improve the customer profile in our proposed model. At first products must be divided into \( k \) groups. For example in movie store we can divide movies in groups using genre and a movie can be member of more than one group. For participation the time in profile we used this formula:

\[ t_r^i = F_t^i * r^i \] (8)

\[ F_t^i = \frac{\sum_{j \in k}(CR_k \cdot r^i - CR_{IT}^i)}{\sqrt{\sum_{j \in k}(CR_k \cdot r^i - CR_{IT}^i)^2 \cdot \sum_{j \in k}(CR_k \cdot T_{now} - CR_{IT}^i)^2}} \] (9)

Where, \( T_{ij} \) is the time that customer \( i \) rated product \( j \) and \( T_{now} \) is the last session time. \( k \) is index of product groups that product \( j \) is member of them and customer \( i \) has interest level of product \( j \) in both period of times \( T_{ij} \) and \( T_{now} \). Here, \( r^i \) represents the element of profile matrix that shows the interest level of customer \( i \) in product \( j \) and \( t_r^i \) represents the element of improved profile matrix. \( CR_k \cdot r^i \) represents the average of interest levels of customer \( i \) in group of products \( k \) at time \( T_{ij} \). and \( CR_{IT}^i \cdot T_{now} \) represents the average of interest levels of customer \( i \) in group of products \( k \) at time \( T_{now} \). Finally, \( CR_{IT}^i \) represents the average of average interest levels of customer \( i \) at time \( T_{ij} \) and \( CR_{IT} \cdot T_{now} \) represents the average of average interest levels of customer \( i \) at time \( T_{now} \).
In our proposed model we used both group preferences and time context to improve customer's profile that never have been used together and we changed the way of improving customer's profile by group preferences.

3.5. Phase 5: Recommendation generation

Product recommendation for customer $i$ is executed as follows. First, similarity between products preferred by customer $i$ and unseen products is computed. This similarity is computed with adjusted cosine formula illustrated in section II. Then, for products that are most similar to preferred products, the interest level of customer $i$ in them is computed. Then, the products are sorted based on calculated interest level. Finally, the top $N$ products would be recommended.

4. Conclusion

Recommender systems are powerful tools that help customers to find products they want to buy from a business. This new technology creates additional value for a business using its purchasing data sets. Recommender systems help customers to find what they really want and save their time and money and help businesses by generating more sales.

In this work, we proposed a new approach for product recommendation based on collaborative filtering. In proposed model with attention to customer profile that is one of the important part of a recommender system, we endeavored to improve the customer profile and recommender system efficiency. In proposed model profile was improved using time context and group preferences. We smoothed the elements of customer profile matrix and realized them. There are some limitations in our study. In future, we must evaluate the proposed model with several data sets and compare with other recommender systems based on collaborative filtering.

5. References

Improving Recommender System's Efficiency Using Time Context and Group Preferences
Mohammad Julashokri, Mohammad Fathian, Mohammad Reza Gholamian, Ahmad Mehrbod
Advances in Information Sciences and Service Sciences. Volume 3, Number 4, May 2011


