Management Science Letters 1 (2011) 449-456

Contents lists available at GrowingScience

Management Science Letters

homepage: www.GrowingScience.com/msl

Improving electronic customers' profile in recommender systems using data mining techniques

Mohammad Julashokri^{a*}, Mohammad Fathian^a, Mohammad Reza Gholamian^a and Ahmad Mehrbod^b

^aDepartment of Industrial Engineering, Iran University of Science & Technology, Tehran, Iran ^bDepartment of Industrial Engineering, Isfehan University of Technology, Esfehan, Iran **ARTICLEINFO ABSTRACT**

Article history: Received March 29, 2011 Received in Revised form June, 18, 2011 Accepted 19 June 2011 Available online 20 June 2011	Recommender systems are tools for realization one to one marketing. Recommender systems are systems, which attract, retain, and develop customers. Recommender systems use several ways to make recommendations. Two ways are using more than the others: collaborative filtering and content-based filtering. In this study, a recommender system model based on collaborative filtering has proposed. Proposed model was endeavored to improve the customer profile in collaborative systems to enhance the recommender system efficiency. This improvement was done using time context and group preferences. Experimental results show that the proposed model has a better recommendation performance than existing models.
Keywords: Recommender systems Customer preference Collaborative filtering Customer profile	
Group preferences	© 2011 Growing Science Ltd. All rights reserved.

1. Introduction

Marketing planning plays an important role on the success of different service providers and businesses. There are various types of marketing strategies such as mass marketing, segmentation and one to one marketing. One to one marketing is an effort to find individual customer's needs and to provide a good response for them. Recommender systems have appeared in e-commerce problems to support product recommendation, which provide one to one marketing (Schafer et al., 2001; Liu & Shih, 2005). Indeed, recommender systems individualize 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. Recommender systems have recently become popular among many well-known e-businesses such as Amazon.com, CDNow.com, Barnes&Noble.com, and MovieFinder.com. There are different versions of recommender systems based on how it is generated such as content-based filtering (CBF) (Cheung et al., 2003; Weng & Liu, 2004; Cho & Kim, 2004; Hung, 2005; Adomavi-cius & Tuzhilin, 2005; Leung et al., 2006; Shih & Liu, 2008), collaborative filtering (CF) (Karypis, 2001; Sarwar et al., 2001; Cheung et al., 2003; Weng & Liu, 2004; Cho & Kim, 2004; Liua 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 * Corresponding author. Tel: +98-E-mail addresses: mohammad.julashokri@gmail.com (M. Julashokri)

© 2011 Growing Science Ltd. All rights reserved. doi: 10.5267/j.msl.2011.06.011 al., 2005; Adomavicius & Tuzhilin, 2005; Choi et al., 2006; Kim et al., 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 works 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. Recommendation methods

There are two well-known types of recommendation methods, which are content-based filtering and collaborative filtering. However, there are some other works based on a combination of these two methods (Resnick & Varian, 1997; Wang et al., 2004). Content-based filtering is proposed using a comparison between content of items and a user profile (Wang & Shao, 2004). In content-based filtering method, characteristics of user's interests are extracted and other items with these characteristics are recommended. In other words, the systems analyze the content of the items and create customer profiles, which represent user's interest in terms of keywords, phrases, and features. Then, the systems compare the user's profile with the content of unknown user to estimate which items could be more interesting to the user (Min & Han, 2005; Park & Chang, 2009).

Collaborative filtering has been more used in ecommerce applications, where it 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 and they are compared to other users to detect similarities among them. Ratings of new items are estimated based on a combination of the ratings of the nearest neighbors to the new items. Nearest neighbors of a user are created based on the correlation of past ratings (Montaner et al., 2003; Park & Chang, 2009). We explain more item-based collaborative filtering methods since our proposed model is based on this method.

2.1.1. Item-based collaborative filtering

Collaborative filtering is one of well-known recommendation method, which has been widely implemented in retail marketing applications. An item-based collaborative filtering recommends products to target customer u based on similarities among items (Sarwar et al., 2001). Item-based collaborative filtering finds similar products 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. In item-based collaborative filtering, 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 are numbers between 1 to 5, where r_{ij} shows 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's (2005) study, elements are binary variables, where r_{ij} is one if the *i*th customer purchases the *j*th product, and zero otherwise. Next step in item-based collaborative filtering is to measure similarity between items, which could be computed in different ways. One common method to compute similarity between items *i* and *j* is adjusted cosine, which is defined as follows (Sarwar et al., 2001),

$$sim(i,j) = \frac{\sum_{u \in U} (R_{u,i} - \bar{R}_u) (R_{u,j} - \bar{R}_u)}{\sqrt{\sum_{u \in U} (R_{u,i} - \bar{R}_u)^2} \sqrt{\sum_{u \in U} (R_{u,j} - \bar{R}_u)^2}},$$
(1)

450

where $R_{u, i}$ and $R_{u, j}$ are elements of profile matrix representing the interest level of *u*th customer in *i*th and *j*th products, respectively. \overline{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.

Next step is to predict the interest level of target customer in products obtained from the last step. This method forecasts the interest level of *u*th user in *i*th item by summing up the ratings given by user on the items, which are similar to *i*. Each ratings must be multiplied according to similarity between items *i* and *j*. The method forecasts interest level of user *u* in product *i* ($P_{u, i}$) as follows (Sarwar et al., 2001),

$$P_{u,i} = \frac{\sum_{all \ similar \ items,N}(s_{i,N} * R_{u,N})}{\sum_{all \ similar \ items,N}(|s_{i,N}|)},\tag{2}$$

where $s_{i, N}$ 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*.

2.2. Customer profile

There are many cases in collaborative filtering method that similarities between user profiles are computed for recommendation generation and a suitable profile increases the precision of recommender system. Lee and Yang (2003) explained that important factor in personalization of researches could be created for each user to predict his/her preferences based on some mathematical methods. Park and Chang (2009) explained that the development of efficient customer profile models is important to improve the recommendation quality of the preferred 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 a user profile structure in their recommender system called HRPCA, which is a matrix of matrices. Their user profile contained users' ratings based 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):

$$r_{ij} = \frac{r_{ij}^c}{r_i^c} + \frac{r_{ij}^b}{r_i^b} + \frac{r_{ij}^p}{r_i^p},$$
(3)

where r_{ij}^c is the total number of clicks of user *i* across product *j* and r_i^c is the total number of clicks of a user *i*. $\frac{r_{ij}^b}{r_i^b}$ and $\frac{r_{ij}^p}{r_i^p}$ are specify the total number of occurrences of basket placements and purchases of user *i* across product *j*, respectively.

In the cases that ratings are directly appointed by user, usually, the precision is low and the 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. Fig. 1 shows the proposed approach generally.



Fig. 1. The proposed model

452 3.1. Phase 1: data preparation

In the first step, the necessary data need to be prepared. There are significant amount of data in realworld applications, which must be cleaned, reorganized, reformatted, etc to use in data mining applications. In this phase, data must be transformed in new appropriate format for application. In our study, we used MovieLens¹ data that is the usual data in recommender systems' test. There are some datasets in MovieLens web site and we use 1 million records. We transformed data from files to SQL tables. Since we did not have the access to a relatively powerful computer to process the data, we used first 100000 record of rating table. Database includes 100 thousand votes from 669 users for 3952 movies. Each customer at least rated 50 movies and we have age, gender and occupation for each customer.

3.2. Phase 2: clustering

In this phase, data are clustered with SOM algorithm. Input fields of algorithm are age, gender, and occupation. SOM clustering determines the number of clusters; then with that specified number, k-means clustering is executed. The cluster number of each customer is used as her/his class label. In the case of our study, SOM algorithm extracted 8 clusters.

3.3. Phase 3: Profile creation

In this step, user profiles is created. In collaborative filtering user's profile usually is a matrix where its rows show users and its columns show items. Elements of user profile matrix show the interest level of a customer to a product. For the case of our study in addition to interest levels, we need the time of voting for using time context. We can insert each rating time to profile matrix because in MovieLens data and each customer's vote to each movie, simultaneously. Fig. 2 shows the profile matrix in our system.



Numbers given in parentheses are periods in terms of months, e.g. 1 means the first month. We used the first 100000 ratings from MovieLens Million data. The data period is 28 months, started at December 2000.

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 clustered in phase 2. Group preferences are used to improve the customer's profile as follows,

$$ir_{ij} = \alpha * r_{ij} + \beta * \bar{r}_{G_ij}, \tag{4}$$

where, r_{ij} represents the element of profile matrix, which shows the interest level of customer *i* in product j and ir_{ij} represents the element of improved profile matrix. \bar{r}_{G_ij} is the average of interest levels of group of customer *i* in product *j*. Weighting values α and β 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 as follows,

$$\alpha + \beta = 1. \tag{6}$$

According to the study of Park and Chang (2009), both α and β are considered as 0.5, but in our proposed model, α is computed by several tests and β is computed using Eq. (6). In our proposed model, we considered several numbers for α and test model with that value and the best value will be considered for α . The other way to improve the customer profile is to use the context. Since customer preferences' may change during time, we used time context to improve the customer profile for our proposed model. 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 a member of more than one group and the parameter of time is considered as follows,

$$ir_{ij} = F_{ij} * r_{ij}, \tag{7}$$

$$F_{ij} = \frac{\sum_{j \in k} \left(CR_{ikT_{ij}} - \overline{CR}_{iT_{ij}} \right) \left(CR_{ikT_{now}} - \overline{CR}_{iT_{now}} \right)}{\sqrt{\sum_{j \in k} \left(CR_{ikT_{ij}} - \overline{CR}_{iT_{ij}} \right)^2 \sum_{j \in k} \left(CR_{ikT_{now}} - \overline{CR}_{iT_{now}} \right)^2}},$$
(8)

where, T_{ij} is the time that customer *i* rates product *j* and T_{now} is the recent time session, *k* is index of product groups that product *j* is a member of them and customer *i* has interest level of product *j* in both period of times T_{ij} and T_{now} . Here, r_{ij} represents the element of profile matrix, which shows the interest level of customer *i* in product *j* and ir_{ij} represents the element of improved profile matrix. $CR_{ikT_{ij}}$ represents the average interest levels of customer *i* in group of products *k* at time T_{ij} and $CR_{ikT_{now}}$ represents the average interest levels of customer *i* in group of products *k* at time T_{now} . Finally, $\overline{CR}_{iT_{ij}}$ represents the average of average interest levels of customer *i* at time T_{ij} and $\overline{CR}_{iT_{now}}$ represents the average of average interest levels of customer *i* at time T_{ij} and $\overline{CR}_{iT_{now}}$ represents the average of average interest levels of customer *i* at time T_{ij} and $\overline{CR}_{iT_{now}}$ represents the average of average interest levels of customer *i* at time T_{ij} and $\overline{CR}_{iT_{now}}$ 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

To recommend a product to customer i, first, similarity among products are preferred by customer i and unseen products are computed. The similarity is computed with adjusted cosine formula illustrated in section II. Then, for products, which are most similar to preferred products, the interest level of customer i is computed. Then, the products are sorted based on calculated interest level. Finally, the top N products are recommended.

4. Experimental results

4.1.Experiment setting

Various experiments are conducted to evaluate the proposed method for product recommendation. The proposed methods are compared to these systems:

- *Simple*: Item-based collaborates filtering with KNN algorithm (k = 30).
- *TimeImprove*: This system is constructed on the basis of simple system, but improved customer profile use time context.
- *StaticGroup*: This system is constructed based on simple system, but improved customer profile using group preferences. Customer groups in this system are static, which means the customers with the same age, gender, and occupation, are in one group. In this system, α is 0.5 for all customers.

454

• *AlphaImprove*: This system is very similar to the proposed model of this paper, but improved customer profile only use group preferences. The value of α is 0.55 for each customer in this system

The MovieLens rating data were divided into a 80% data set for training and a 20% data set for testing to verify the quality of the recommendations. We constructed customer profile matrix by training set and generated 10 recommendations for each customer. Then we calculated evaluation metrics by generating recommendations and test set data to evaluate the proposed recommender system's efficiency.

4.2. Evaluation metrics

Various metrics have been used to evaluate the recommender system's accuracy. In this paper, we use two general metrics, precision and recall, where they are commonly used to measure the quality of a recommendation (Cho & Kim, 2004; Karypis, 2001; Sarwar, et al., 2001; Shih and Liu, 2008; Albadvi and Shahbazi, 2009). Recall is defined as the ratio of the number of products recommended correctly, products in both recommendation set and test set, and it is set using the following,

$$recall = \frac{number of correctly recommended products}{number of test set products}.$$
(9)

Precision is defined as the ratio of number of products recommended correctly to number of products in recommendation set,

$$precision = \frac{number \ of \ correctly \ recommended \ products}{number \ of \ recommendation \ set \ products}.$$
(10)

These measures are easy to compute, but as the size of the recommendation increases, recall also increases, which increases the precision of our computations. Therefore, a well-known combination metric called F1 metric is produced to provide a balance trade-off between recall and precision (Cho & Kim, 2004; Karypis, 2001; Sarwar, et al., 2001; Shih & Liu, 2008; Albadvi & Shahbazi, 2009) as follows,

$$F1 = \frac{2 \times recall \times precision}{recall + precision}.$$
(11)

However, an increase on F1 metric yields higher quality results, which increases recommender system's efficiency. Each metric is measured for each customer, and the average value is computed to measure the quality of recommendations.

4.3. Evaluation results

This section presents a detailed experimental evaluation of different recommender systems described in section 4.1. This experiment verifies that the proposed method is relatively a good alternative method to enhance the quality of recommender system. To evaluate the proposed model we performed different experiments. The rest of this section illustrates the experiments.

4.3.1. Calculating α

Calculating the best value for α is one of the important factors in our proposed model. In this step, a recommender system using the proposed model was generated. Similarity equation, neighborhood size, and the number of recommendations have changed to determine the best recommender system's efficiency. First, different values of neighborhood size and the number of recommendations have tested for Pearson, adjusted cosine, and cosine similarity equations and the results indicated that cosine equation is the best one for our dataset. In the next step, different values of neighborhood size and cosine similarity equation have been tested for different values of neighborhood size and cosine similarity equation have been tested for different values of number of recommendations. The results also indicated that 10 recommendations are the best for our dataset. In the last step,

different values of neighborhood size have tested with 10 recommendations and cosine similarity equation. The results shown 10 neighbors are the best for our dataset.





Fig. 3. Comparison between different α values



The proposed model was built with 10 recommendations, 10 neighbors, and cosine similarity equations. Different values between 0 and 1 were tested for α by proposed recommender system. Fig. 3 shows the results. As shown Fig. 3, the proposed model with $\alpha = 0.55$ has the best efficiency compared with tested models.

4.3.2. Evaluation of different systems

In this experiment, we compared the performance of the different systems described in section 4.1 with the performance of the proposed model. Fig. 4 shows F1 metric for different recommender systems. As shown in Fig. 5, the proposed model performs well compared with other alternative methods. This experiment shows that profile improvement via time context and group preferences can improve the recommender system's quality.

5. Conclusion

Recommender systems are powerful tools that help customers find products they want to buy from a business. This new technology generates additional value for a business using its buying data sets. Recommender systems help customers find what they really need 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's profile, which is one of the important parts 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.

References

- Adomavicius, G., & Tuzhilin, A. (2005). Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. *IEEE Transactions on Knowledge and Data Engineering*, 17(6), 734-749.
- Albadvi, A., & Shahbazi, M. (2009). A hybrid recommendation technique based on product category attributes. *Expert Systems with Applications*, 36(9), 11480 -11488.
- Boucher-Ryan, P. D., & Bridge, D. (2006). Collaborative recommending using formal concept analysis. *Knowledge-based Systems*, 19(5), 309 -315.

Bult, J. R., & Wansbeek, T. J. (1995). Optimal selection for direct mail. *Marketing Science*, 14(4), 378-394.

Burke, R. (2002). Hybrid recommender systems: Survey and experiments. User Modeling and User-adapted Interaction, 12(4), 331-370.

456

- Cheung, K.-W., Kwok, J. T., Law, M. H., & Tsui, K.-Ch. (2003). Mining customer product ratings for personalized marketing. *Decision Support Systems*, 35(2), 231-243.
- Cho, Y. H., & Kim, J. K. (2004). Application of Web usage mining and product taxonomy to collaborative recommendations in e-commerce. *Expert Systems with Applications*, 26(2), 233-246.
- Choi, S. H., Kang, S., & Jeon, J. Y. (2006). Personalized recommendation system based on product specification values. *Expert Systems with Applications*, 31(3), 607-616.
- Hung, L. P. (2005). A personalized recommendation system based on product taxonomy for one-to-one marketing online. *Expert Systems with Applications*, 29(2), 383 -392.
- Hwang, C. S. (2010). Genetic algorithms for feature weighting in multi-criteria recommender systems. *Journal* of Convergence Information Technology, 5(8), 126-136.
- Karypis, G. (2001). Evaluation of item-based top-n recommendation algorithms. In Proceedings of the ACM CIKM conference, 247-254.
- Kim, B. M., Li, Q., Park, C. S., Kim, S. G., & Kim, J. Y. (2006). A new approach for combining content-based and collaborative filters. *Journal of Intelligent Information Systems*, 27(1), 79-91.
- Lee, C. H., Kim, Y. H., & Rhee, P. K. (2001). Web personalization expert with combining collaborative filtering and association rule mining technique. *Expert Systems with Application*, 21(3), 131-137.
- Lee, W. P., & Yang, T. H. (2003). Personalizing information appliances: A multi-agent framework for TV programme recommendation. *Expert Systems with Applications*, 25(3), 331-341.
- Leung, C. W.-K., Chan, S.C.-F., Chung, F.-L. (2006). A collaborative filtering framework based on fuzzy association rules and multiple-level similarity. *Knowledge and Information Systems*, 10(3), 357-381.
- Lihua, W., Lu, L., Jing, L., & Zongyong, L. (2005). Modeling user multiple interests by an improved GCS approach. *Expert Systems with Applications*, 29(4), 757-767.
- Liu, D.-R., & Shih, Y.-Y. (2005). Hybrid approaches to product recommendation based on customer life time value and purchase preferences. *The Journal of Systems and Software*, 77(2), 181-191.
- Liu, D.-R., & Shih, Y.-Y. (2005). Integrating AHP and data mining for product recommendation based on customer lifetime value. *Information & Management*, 42(3), 387-400.
- Min, S. H., & Han, I. (2005). Detection of the customer time-variant pattern for improving recommender systems. *Expert Systems with Application*, 28(2), 189-199.
- Montaner, M., Lopez, B., & De La Rosa, J. L. (2003). A taxonomy of recommender agents on the Internet. *Artificial Intelligence Review*, 19(4), 285-330.
- Park, Y.-J., & Chang, K.-N. (2009). Individual and group behavior-based customer profile model for personalized product recommendation. *Expert Systems with Applications*, 36(2), 1932 -1939.
- Resnick, P., & Varian, H. R. (1997). Recommender systems. Communication soft the ACM, 40(3), 56-60.
- Sarwar, B., Karypis, G., Konstan, J., & Riedl, J. (2001). Item-based collaborative filtering recommendation algorithms. In Proceedings of the international world wide web conference, Hong Kong, 285-295.
- Sarwar, B., Karypis, G., Konstan, J., & Riedl, J. (2000). Analysis of recommendation algorithms for ecommerce. In Proceedings of the ACM Conference (ElectronicCommerce), 158-167.
- Schafer, J. B., Konstan, J. A., & Riedl, J. (2001). E-commerce recommendation applications. *Journal of Data Mining and Knowledge Discovery*, 5(1/2), 115-152.
- Semeraro, G., Lops, P., & Degemmis, M. (2005). Word Net-based user profiles for neighborhood formation in hybrid recommender systems. In Proceedings of the 5th international conference on hybrid intelligent systems (HIS'05), 291-296.
- Shih, Y. Y., & Liu, D. R. (2008). Product recommendation approaches: Collaborative filtering via customer life time value and customer demands. *Expert Systems with Applications*, 35(1–2), 350-360.
- Wang, F. H., & Shao, H. M. (2004). Effective personalized recommendation based on time-framed navigation clustering and association mining. *Expert Systems with Application*, 27(3), 365-377.
- Wang, Y. F., Chuang, Y. L., Hsu, M. H., & Keh, H. C. (2004). A personalized recommender system for the cosmetic business. *Expert Systems with Applications*, 26(3), 427-434.
- Weng, S. S., & Liu, M. j. (2004). Feature-based recommendations for one-to-one marketing. *Expert Systems with Applications*, 26(4), 493-508.
- Yang, H.-L., & Yang, H.-F. (2011). Recommendation mechanism based on multi-attribute utility theory. *International Journal of Digital Content Technology and its Applications*, 5(3), 373 -382.