An Implicit Rating based Product Recommendation System

Chieh-Yuan Tsai†
Department of Industrial Engineering and Management
Yuan Ze University, Taoyuan City, Taiwan
†Tel: (+886) 3-463-8800 ext. 2512, Email: cytsai@saturn.yzu.edu.tw

Sih-Wei Shen
Department of Industrial Engineering and Management
Yuan Ze University, Taoyuan City, Taiwan

Abstract. Collaborative filtering (CF) is one of the most popular approaches in recommendation systems. Typically, collaborative filtering looks for users who share the same rating patterns with the active user and uses the ratings to calculate a prediction for the active user. However, explicit ratings are not always available in many physical stores. To solve the problem, this paper develops a recommendation system that can derive user preference ratings from users' purchase history without any explicit feedback provided by user. In addition, this research also considers the time interval between purchased time and current time (recency) and item relationship. Finally, max-min fuzzy theory is used to combine the two factors. Through experiments, the proposed approach shows much better performance than the system which considers only the factor of item purchase frequency. This study also shows that the performance of the proposed system outperforms the one without considering the time interval between purchased time and current time as well as item relationship.

Keywords: recommendation systems, implicit feedback, collaborative filtering, max-min fuzzy.

1. INTRODUCTION

In the information overloaded age, users usually hope to find what they want without wasting much time. To solve this burden, recommendation systems have been emerged in response to this problem. Fundamental knowledge and techniques for developing recommendation systems have been proposed in recent decades, including content-based filtering (Pazzani and Billsus 1997), collaborative filtering (Yu et al. 2004, Konstan et al. 1997) and hybrid approaches (Balabanovic and Shoham 1998, Salter and Antonopoulos 2006, Wei et al. 2008).

Collaborative filtering relies on users whose preferences (or interesting) are similar to those of target users and recommends items that users have liked. Because collaborative filtering depends on user behavior, it attracted much attentions resulting in significant progress and being adopted by some successful commercial systems, including Amazon (Linden et al. 2003) and Netflix (Bennet and Lanning 2007). In many collaborative filtering systems, user behavior can be derived from user's interaction with the system which is called feedback. Feedback can be divided into explicit feedback and implicit feedback. Explicit feedback refers to the relevant information provided by users directly. The typical explicit feedback is the user rating on items (Roy et al. 2010). However, explicit feedback is not always available in practice. Thus, some recommendation systems discuss implicit feedback issues.

Research in implicit feedback can be divided into two parts. One focuses on using demographic data such as age, education, income or gender to find a set of users similar to the target user and calculate the rating by the set of similar users (Zou et al. 2009, Liu and Shih 2005, Liu et al. 2009, Wang and Zhou 2012). For example, Zou et al. (2009) used demographic data to find similar users and combined other information such as how many times users visited museum and how long they stayed in museum to recommender. Wang and Zhou (2012) utilized demographic data to define rating of different attributes and calculate the rating by what attributes of a target user has. Although these researches use demographic data to find similar group of users, they do not take user's real transaction data into consideration and still require some explicit feedback.

Another group of researches in implicit feedback focuses on how transform user preference rating from purchase history, browsing history, search patterns, or even mouse movements (Lee et al. 2010, Kim et al. 2005, Choi et al., 2012). For example, Lee et al. (2010) proposed a pseudo rating matrix which definite rating by items lunch time and users purchase time. Choi et al. (2012) proposed an equation to transform user rating by how many times user buys an item. However, these researches don't consider the difference of time interval between purchased time and current time.

To solve the above problems, the purpose of this paper is to develop a recommendation system to derive user preference ratings from users' purchase history without using explicit feedback provided by the user. This recommendation system takes purchase frequency, purchase cycle time and purchase time-interval to established user preference rating. Moreover, this research also considers the time interval between purchased time and current time and item relationship. Finally, this research uses max-min fuzzy theory to combine these factors.

2. RESEARCH METHOD

The framework of the proposed implicit rating based product recommendation system consists of three main steps as shown in Figure 1. They are (1) constructing user preference rating using purchase frequency, cycle time and purchase time-interval; (2) predicting user preference rating using best-n-neighbors; (3) generating an item recommendation list through neighborhood of the target user with considerations of purchase recency and item association.

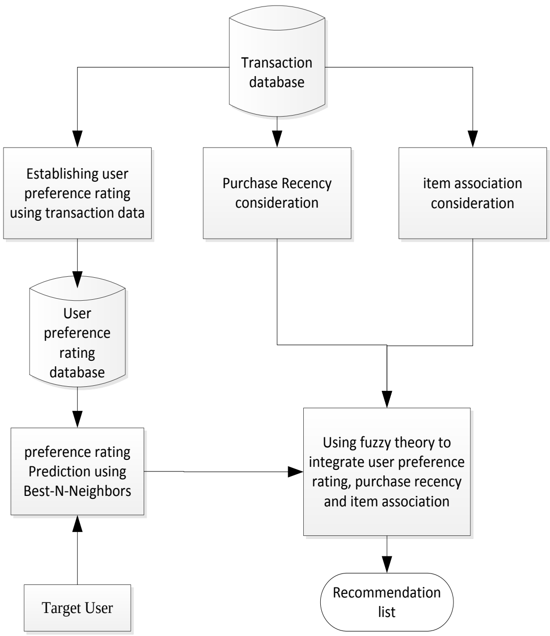

Figure 1: The framework of the proposed implicit rating recommendation system.

2.1 Establishing User Preference Rating using Transaction Data

As mentioned before, it is quite often that user preference rating is difficult to obtain through explicit feedback. Therefore, this research constructs user preference rating using transaction data. The preference rating construction is based on three observations. First, if the purchase frequency (count) for an item is high, the preference rating for the item will be high. Second, if the cycle time between two purchases for the same item is short, the preference rating for the item will be high. Third, if the time-interval between transaction time and current time is short, the preference rating for the item should be high.

Let $T_{u,i,t}$ represent tth transaction time that user u purchases item i. $CT_{u,i,t}$ is the cycle time that that user u buy item i between tth and t+1th transactions. It can be evaluated as:

$$CT_{u,i,t} = T_{u,i,t+1} - T_{u,i,t} \tag{1}$$

Let $TT_{u,i,t}$ represent time-interval between t+1th transaction time and current time T_{now} for user u and item i. $TT_{u,i,t}$ can be evaluated as:

$$TT_{u,i,t} = T_{now} - T_{u,i,t+1} \tag{2}$$

To construct user preference rating of user u for item i, $PD_{u,i}$, the following equation is formulated:

$$PD_{u,i} = \sum_t \left(\frac{MinCT_i}{CT_{u,i,t}} \times \frac{1}{TT_{u,i,t}}\right) \tag{3}$$

$MinCT_i$ is the minimum cycle time for all users who ever purchased item i, which can be derived by:

$$MinCT_i = \min_{all\ u, all\ t} \{CT_{u,i,t}\} \tag{4}$$

2.2 Preference Rating Prediction using Best-N-Neighbors

Typically, a collaborative filtering recommendation system predicts the rating of an item for a target customer based on how similar customers rated the same item previously. In this research, we use Equation (7) to calculate the similarities between target user u and other user v:

$$SIM_{u,v} = \frac{\sum_{i \in M_{u,v;}} \left|(PD_{u,i} - \bar{P}_{u,v}) \times (PD_{v,i} - \bar{P}_{v,u})\right|}{\sqrt{\sum_{i \in M_{u,v;}} (PD_{u,i} - \bar{P}_{u,v})^2} \sqrt{\sum_{i \in M_{u,v;}} (PD_{v,i} - \bar{P}_{v,u})^2}} \tag{7}$$

$$\bar{P}_{u,v} = \frac{\sum_{i \in M_{u,v}} PD_{u,i}}{|M_{u,v}|} \tag{8}$$

here $\bar{P}_{u,v}$ is the average user preference rating for user u in which items exist in both user u and user v, and $M_{u,v}$ is the set of items which exist in transactions made by user *u* and user *v*.

After obtaining $SIM_{u,v}$ for all user *v*, a set of neighbors similar to the target user *u* will be retrieved. In this research, the best-*k*-neighbor technique is adopted. That is, top-*k* users who have higher similarity with the target user are considered as neighbors.The predicted preference rating for item *i* by the target user *u*, $R_{u,i}$ can be estimated by the follow equation:

$$R_{u,i} = \frac{\sum_{v\in N} PD_{v,i} \times SIM_{u,v}}{\sum_{v\in N} \left| SIM_{u,v} \right|} \tag{9}$$

2.3 Recency Consideration

It's clear that user's desire for an item will reduce if the time between current time and last purchase time is long. Similarly, if the time between current time and last purchase time for an item is too short, the purchase desire for the item will reduce also. Therefore, the predicted preference rating for an item should be adjusted according to the above idea. Let $WT_{u,i}$ be the adjusted weight for user u and item i and can be defined as:

$$WT_{u,i} = (\frac{RT_{u,i}}{\overline{CT}_{u,i}})^{0.5} \quad if \quad RT_{u,i} \le \overline{CT}_{u,i} \tag{10}$$

$$WT_{u,i} = (\frac{\overline{CT}_{u,i}}{RT_{u,i}})^{0.5} \quad if \quad RT_{u,i} > \overline{CT}_{u,i} \tag{11}$$

where $RT_{u,i}$ is time between current time and the last time of purchasing item *i* for user *u*, also called recency for item *i* and user u. $\overline{CT}_{u,i}$ is the average cycle time of buying item *i* for user *u*. If user *u* did not buy item *i* more than one time, we will take $\overline{CT}_{v,i}$ of user *u*'s neighbor *v* as $\overline{CT}_{u,i}$.

2.4 Item Association

Some items are often purchased together. Therefore, item association should be involved when we generate item suggestion list. Let Supp({*i*, *i'*}) represent how many times item *i* and item *i'* happen in same transaction data. *Int*({*i*, *i'*}) represents two items intimacy which can be defined as:

$$Int(\{i,i'\}) = \frac{Supp(\{i,i'\})}{Supp(\{i\}) + Supp(\{i'\}) - Supp(\{i,i'\})} \tag{12}$$

2.5 Fuzzy Association for Information Integration

This research uses fuzzy theory to integrate two items intimacy, user preference, and the time between current time and last purchase time (recency). Max-min fuzzy theory is based on the ratio of the number of all maximum fair dominating sub vectors to the number of all possible sub vectors. It is shown how this definition extends the maximum fairness relation, how it helps to solve the problems with maximum fairness, and how it numerically emphasises fairness. Max-min fuzzy theory fairness can be a formally efficient definition of a fairness concept.

Let $F_{u,i}$ represent the strength preference that user *u* will buy item *i*, which can be derived by

$$F_{u,i} = R_{u,i} \times WT_{u,i} \tag{13}$$

where $R_{u,i}$ is the predicted preference rating for item *i* by the target user *u* and $WT_{u,i}$ is the adjusted weight for user *u* and item *i*. Next, min-max normalization is applied to $F_{u,i}$ so that modified strength $F'_{u,i}$ will be between 0 and 1. Third, this research uses the max-min composition proposed by Zadeh (1965) to generate an recommendation list for the target user. Let $\vec{a}_u$ be the vector of modified strength $F'_{u,i}$, which is defined as:

$$\vec{a}_u = [F'_{u,A}, F'_{u,B} ... F'_{u,i}] \tag{14}$$

Let the set of the two item intimacy in Equation (12) be a fuzzy relation, *IR*. That is,

$$\mathrm{IR} = \left\{ (\mathrm{x}, y), \mathrm{Int}(\{\mathrm{x}, \mathrm{y}\}) \middle| (x, y) \in i \times i \right\} \tag{15}$$

Specifically, the proposed fuzzy association $\underline{\mathrm{FA}}[\vec{a}_u, IR)]$ can be defined as:

$$\underline{\mathrm{FA}}[\vec{a}_u, IR] \rightarrow \vec{a}_u^{\,new} \tag{16}$$

where $\vec{a}_u^{\,new}$ is the resulting vector after conducting fuzzy association. That is, the new vector $\vec{a}^{new}$ can be obtained by:

$$\vec{a}_u^{\,new} = [FTI_{u,A}, FTI_{u,B} .., FTI_{u,i}] \tag{17}$$

where

$$FTI_{u,i} = \vee_i [F'_{u,i} \wedge Int(\{x,y\})] = \max_{all\ i} \min[F'_{u,i}, Int(\{x,y\})]$$

, $\vee$ and $\wedge$ represent the fuzzy max and fuzzy min respectively.

3. EXPERIMENTAL ILLUSTRATION

To test the feasibility of the proposed implicit rating based product recommendation system, a simple shopping store is used as an example. It is assumed that there are six user types in which each user type has unique shopping patterns. Table 1 summarizes shopping patterns of each user type. For example, customers with type I visits the shopping store within five to nine days and purchases no greater than eight items in each visit.

Table 1: The shopping patterns for six user types.

User Type	Interval arrival time (day)	Maximum number of purchases
I	[5, 9]	8
II	[6, 10]	8
III	[9, 13]	13
IV	[10, 14]	13
V	[14, 18]	18
VI	[15, 19]	18

In this shopping store, there are 120 product items available where the average purchase cycle time for each item is classified as short term, mid-term or long-term. The average purchase cycle time for short-term items such as like cookies, fruit and fresh food is around two weeks. The average purchase cycle time for mid-term items such as microwave food and instant noodles is around one month. The average purchase cycle time for long-term items such as like shampoo and toilet paper is over one month. In addition, different user types have different shopping patterns. For example, users of type V tends to buy baby diapers in which users of other types seldom buy this. Moreover, the average purchase cycle time for items for different user types might be different.

3.1 Case Illustration

The proposed implicit rating based product recommendation system is implemented using C# and tested on a PC with Core i7 2.40GHz and 8GB memory. In this illustration, the total user number is 300; the number of neighbors is 25, and the number of recommended items is 5. The recommendation system will re-evaluate every simulation day. For example, if the simulation day is day 741, the recommendation system will consider day 365 to day 740 as transaction data. It means the recommendation system will update preference matrix every day and provide users updated recommendation list.

In order to establish user preference, the system will read transaction data first and then establish user preference. This research constructs user preference rating using transaction data. Through Equation (3), user preferences can be derived. Next, the similarities between user *u* and user *v* can be calculated using Equation (7). Notes that the equation only considers the average user preference rating of which items both exist in user *u* and user *v*. According to Equation (9), the predicted preference rating for each item can be calculated.

To understand the performance of the proposed system, four different Models are compared. Choi et.al (2012) proposed a recommended method computed solely based on the purchase data of user *u*. In Choi et al. (2012), the preference rating for user *u* and item *i* , called Model 0, is defined as how many times user *u* have purchased item *i* divided by how many times item *i* have be purchase by all users. The recommended result using the predicted preference rating generated by Equation (9) is called Model 1. The recommended result using the strength preference rating generated by Equation (13) is called Model 2. The recommended result using the integrated preference rating generated by Equation (16) is called Model 3.

For example, if user U74 visits shopping store at day 731. The transaction record shows that U74 purchased 7 items of I7, I13, I19, I25, I37, I118 and I117 at that day. If the number of recommended items is set as 10, for the U74 the number of correct recommended items for Model 0 is 5, Model 1 is 6, Model 2 is 6 and Model 3 is 7. Similarly, if U241 visits shopping store at day 731 and he/she purchased 12 items as I1, I2, I11, I17, I23, I29, I35, I41, I53, I59, I96, I98 and I116. The number of correct recommended items for U241 is 6 for Model 0, 8 for Model 1, 9 for Model 2, and 9 for Model 3. Table 2 shows the different model's recommended lists for U74 and U241 at day 731.

Table 2: A set of recommended lists for user U74 and U241.

Day	User	Model type	Recommended items	No. of correct items
731	U74	Model 0	I111, I113, I117, I118, I99, I115, I37, I19, I31,I112	5
		Model 1	I13, I111, I25, I117, I118, I37, I49, I97, I19, I31	6
		Model 2	I13, I19, I25, I97, I117, I118, I31, I37, I49, I113	6
		Model 3	I7, I13, I19, I25, I117, I118, I31, I37, I49, I97	7
731	U241	Model 0	I1, I11, I47, I33, I29, I35, I114, I96, I98, I93	6
		Model 1	I1, I11, I53, I17, I48, I33, I29, I2, I35, I41	8
		Model 2	I1, I2, I11, I17, I23, I29, I35, I41, I48, I53	9
		Model 3	I1, I2, I11, I23, I29, I35, I41, I53, I17, I48	9

3.2 Performance Comparison

In order to understand the performance of the proposed method, recall, precision, and F-score is used to evaluate the quality of recommendations. Recall is the fraction of the correct recommended items divided by the total number of recommended items by a recommendation method. Precision is the fraction of correct recommended items divided by the total number of items that a user purchased. F-Score combines precision and recall to evaluate the quality of a recommender.

In this experiment, the number of customers is 300 and number of neighbors is 25, and the number of recommended items is 5. Since the simulation program contains random procedure, five different datasets are generated and tested for different Models. Table 3 shows the variance among the five datasets is relatively small in all models. Through Table 3, we can find the five data F-score variance by all models are all small than 0.00004. Because the variance are so small, that we assume there are no significant differences between data 1 to data 5. In the following discussion, all experiments will be conducted five time and take the average of the five experiments as the final value.

Table 3: The average of F-scores under different number of users.

Model Type	Sum	Mean	Variance
Model 0	2.9282644	0.585653	0.000002
Model 1	3.3718301	0.674366	0.000016
Model 2	3.7806534	0.756131	0.000037
Model 3	3.8353886	0.767078	0.000013

4. CONCLUSIONS

In the information overloaded age, users usually want to find what they want without wasting much time. To solve this burden, recommendation systems have emerged in response to this problem. Among them, collaborative filtering approach is one of the most popular ones. However, many collaborative filtering recommendation systems use explicit feedback (ratings) which are collected directly from users to infer the possible recommendation. However, in many cases, explicit feedback is hard to obtain and not always available in practice. To solve the above problems, this paper develops a recommendation system to derive user preference ratings from users' purchase history without any explicit feedback provided by the user. This recommendation system takes purchase frequency, purchase cycle time and purchase time-interval to established user preference rating. In addition, this study uses max-min fuzzy theory to combine time weight and item association to get more accurate conclusion.

Some possible extensions are summarized as follows. This research is based on simulation data. Therefore, testing the proposed system in a practical shopping store should be helpful. In practical situation, shopping store might give different discount to different items. The sale price might be changed and become a significant factor affecting user purchase behavior. In the future, sale price can be taken into consideration when making product suggestions.The proposed recommendation system is designed for benefiting customers. It might be interesting if further research can take company profit in to consideration when making suggestion. Currently, the special events and seasonal attributes are not taken into account. For example, ice cream is usually purchased in summer, ingredients for hot pot are usually sold in winter, and candy or toys will popular before special holiday. The future work can consider the influences of time periods.

REFERENCES

Balabanovic, M. and Shoham, Y. (1998) Content-based, collaborative recommendation. *Communications of the ACM*, **40(3)**, 66-72.

Bennett, J. and Lanning, S. (2007) The Netflix Prize, *Proceedings of KDD cup and workshop*.

Choi, K., Yoo, D., Kim, G., and Suh, Y. (2012) A hybrid online-product recommendation system: Combining implicit rating-based collaborative filtering and sequential pattern analysis. *Electronic Commerce Research and Applications*, **11(4)**, 309-317.

Konstan, J., Miller, B., Maltz, D., Herlocker, J., Gordon, L., and Riedl, J. (1997) GroupLens: Applying collaborative filtering to usenet news. *Communications of the ACM*, **40(3)**, 77-87.

Kim, Y. S., Yum, B. J., Song, J., and Kim, S. M. (2005) Development of a recommender system based on navigational and behavioral patterns of customers in e-commerce sites. *Expert Systems with Applications*, **28**, 381-393.

Linden, G., Smith, B., and York, J. (2003) Amazon.com recommendations: item-to-item collaborative filtering. *IEEE Internet Computing*, **7**, 76–80.

Liu, D. R. and 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**, 181-191.

Liu, D. R., Lai, C. H., and Lee, W. J., (2009) A hybrid of sequential rules and collaborative filtering for product recommendation. *Information Sciences*, **179(20)**, 3505-3519.

Lee, S. K., Cho, Y. H., and Kim, S. H., (2010) Collaborative filtering with ordinal scale-based implicit ratings for mobile music recommendations. *Information Sciences*, **180(11)**, 2142-2155.

Pazzani, M. and Billsus, D. (1997) Learning and revising user profile: the identification of interesting web sites. *Machine Learning*, **27(3)**, 313-331.

Roy, S. B., Sihem, A. Y., Chawla, A., Das, G., and Yu, C. (2010) Space efficiency in group recommendation. *The VLDB Journal*, **19(6)**, 877–900.

Salter, J. and Antonopoulos, N. (2006) Cinema screen recommender agent: combining collaborative and content-based filtering. *IEEE Intelligent Systems*, **21(1)**, 35-41.

Wang, X. and Zhou C. (2012) A collaborative filtering recommendation algorithm using user implicit demographic information. *Computer Science & Education*, **7**, 935-939.

Wei, C. P., Yang, C. S., and Hsiao, H. W. (2008) A collaborative filtering-based approach to personalized document clustering. *Decision Support Systems*, **45(3)**, 413-428.

Yu, K., Schwaighofer, A., Tresp, V., and Xiaowei, X. (2004) Probabilistic memory-based collaborative filtering. *Knowledge and Data Engineering*, **16(1)**, 56-69.

Zhou, S., Zhou, X., Yu, Z., Wang, K., Wang, H., and Ni, H. (2009) A recommendation framework towards personalized services in intelligent museum. *International Conference on Computational Science and Engineering*, 229-234.