Hybrid Approach for product Recommendations using Collaborative filtering

Chinmay Puranik[1], Grantej Otari[2]
P.G. Student, Department of Computer Science & Engineering, KIT's College of Engineering, Maharashtra, India[1]
Assistant Professor, Computer Science & Engineering, KIT's College of Engineering, Maharashtra, India[2]

ABSTRACT: In the rapidly growing mobile music market the promising solution is Collaborative filtering (CF)-based recommender systems. In the mobile web environment, a traditional CF system that uses explicit ratings to collect user preferences has a limitation: because of poor interfaces and high telecommunication costs, mobile customers find it difficult to rate their tastes directly. For the mobile web, implicit ratings are more desirable but commonly used cardinal scales for representing preferences are also unsatisfactory because they may increase estimation errors. The proposed methodology based on both implicit rating and less ambitious ordinal scales. As an implicit rating approach, a mobile Web usage mining (mWUM) technique is suggested, and a specific consensus model is employed to generate an ordinal scale-based customer profile. In the mobile Web environment, an experiment with the participation of real mobile Web customers shows that the proposed methodology provides better performance than existing CF algorithms.

KEYWORDS: **Collaborative filtering (CF), Implicit Ratings, Recommendations;**

I. INTRODUCTION

Collaborative filtering is a method of making automatic predictions (filtering) about the interests of a user by collecting preferences or taste information from many users. Collaborative filtering methods have been applied to many different kinds of data including sensing and monitoring data. It is one of the most successful recommendation techniques. The basic idea is to recommend items to a target user by predicting their utility for that user through previous ratings by other users.

In the mobile Web environment, an increasing variety of content has been made available with the mobile music market witnessing a particularly rapid growth. However, customers still experience a great deal of frustration when searching for the music they want on mobile Web devices, owing to the inefficiencies of searching sequentially. When a customer uses a cell phone to log on to a site to download music, the site presents a list of the available best-selling or newest music. The customer pages through the list and selects items to inspect. If the customer likes the item, he/she may buy it. Otherwise, the customer repeats these steps until he/she stumbles over the right item or gives up. With this method, the number of items the customer would be expected to encounter before finding the desired item could far exceed the acceptable level. These difficulties are partly attributable to typical cell phone features. Compared to PCs, cell phones have smaller screens, fewer input keys, and less sophisticated browsers. Thus, the user interface of the mobile Web application is not as friendly as that of typical Web applications. To make searching more acceptable, a more efficient search aid suggesting only the items meeting the customer's preference is therefore necessary.

To identify preferences and make recommendations .CF-based recommender systems require a customer profile. To create a customer profile, two profiling techniques are commonly used: explicit ratings and implicit ratings. Explicit ratings are techniques in which customers examine each item and assign it a value on an agreed rating scale. Implicit ratings are techniques that gather information about a customer's shopping behaviors and represent preferences as ratings without the customer's intervention.

II. RELATED WORK

Cook (2006) and Kress (1985) [5] suggested a consensus method to represent the intensity of preference (called the CK method). Unlike previous distance-based methods, the CK method attempts to express the strength of preference by diversifying the preference specifications of Kemeny and Snell's method. Specifically, voters are permitted to express the intensity of their preferences.

Cho and Kim (2004) [3] suggested a formal way of using a WUM technique to ascertain implicit ratings in their research on a shopping-mall's product recommendations. Kim (2004) suggested a basic method of mWUM in their research on mobile image recommendations. Their study proved that WUM in the mobile environment is an effective technique in alleviating the sparsity problem and in enhancing the quality of the recommendation. However, both of these studies used cardinal scales for ratings and did not investigate the possibility that an ordinal scale might overcome some of the limitations of cardinal scales when used with the implicit rating approach.

Aggarwal and Yu (2000) [1] described the tracking of user behavior on Web sites to implicitly identify user preferences. This paper discusses an overview of data mining techniques for personalization. The idea in personalization in web business applications is to collect information about the customer in some direct or indirect form and use this information in order to develop personalization tools. Thus, an electronic commerce site may continue to keep information about the behavior of the customer in some explicit or implicit form.

Sarwar (2000) [6] suggested a method of using SVD for matrix factorization, which provides the best lower rank approximations of the original matrix. Instead of identifying a neighbourhood of similar users, the item-to-item correlation approach analyzes the user-item matrix to identify relationships between different items and uses these relations to compute the prediction score for a given user-item pair.

Claypool and Pazzani (1999) [9] implemented separate collaborative and content-based systems and combined the outputs (ratings) obtained from the individual systems into one final recommendation by using either a linear combination of ratings or a voting scheme. Fab is based on traditional CF, but it also maintains content-based profiles for each user. These hybrid recommendation approaches have also been used to effectively overcome both new user and new item problems.

III. SYSTEM OVERVIEW

Proposed system is going to implement collaborative filtering in order to cater the mobile user downloading music.

General behaviour patterns in the mobile Web

System identifies the customer's preference from information on his/her previous behaviours, and then it creates the customer profile based on these preferences. The following general behaviour patterns in the mobile Web environment are defined:

Ignore: Not clicking on the title of music on the list page and moving on to the next music title or pages.

Click-through: Clicking on a certain title on the list page and viewing the detailed information.

Pre-listen: Pre-listening a sample of the music that was selected.

Download: Downloading the music that has been clicked-through or pre-listened.

Based on these patterns, all music items are classified into four groups as downloaded music, pre-listened music, clicked-through music, and other music items (ignored). This classification provides an is-a relation between the different groups such that downloaded music is-a music pre-listened and music pre-listened is-a music clicked-through. From this relation, it is reasonable to obtain a preference order between music items such that {music ignored (never clicked)} < {music clicked-through} < {music pre-listened} < {music downloaded}.

	Music					
Session	m_1	m_2	m_3	m_4	m_5	m_6
SA	3	2	2	0	0	1
SB	2	3	2	0	1	1
SC	2	0	3	1	2	2
SD	1	0	0	2	3	1
SE	2	0	1	2	0	2

Phase 1: mobile Web usage mining (mWUM)

Step 1.1. Data pre-processing

Data pre-processing in mWUM consists of a series of dependent tasks including data cleaning, user identification, and session identification. It is similar to WUM in the typical Web environment, but it is simpler and more advantageous because of the unique characteristics of the mobile Web.

First, in the typical Web environment, most Web pages contain numerous irrelevant items such as gif, jpg, and swf, and removing them is a time-consuming job. In the mobile Web, however, most pages consist mainly of text, and pages are small with limited components. Second, in the mobile Web environment, all customers must subscribe to network carriers and receive a unique cell-phone or equipment number (EN). This prerequisite user registration prevents confusion in the user identification task. Furthermore, it provides the opportunity to utilize a customer's demographic and social information for better recommendations. Third, in the mobile Web environment, all pages accessed by a customer at a certain time are recorded with identical values in the specific field (e.g., SESSION_ID) of the log, and this SESSION_ID field can be successfully used to identify each individual session.

System creates session files by sorting all log entries according to the SESSION_ID's values and grouping them together for each customer.

Step 1.2. Customer behaviour mining

Once the data pre-processing tasks have been completed in Step 1, a customer's session file is created. In this step, a specific matrix called the customer action set is created by using the session file. The customer action set C=(Csj) is a matrix containing the numerical weights of the target customer's shopping behaviours for music items. It is defined as follows:

mWUM Algorithm:
INPUT: customer transactions in raw mobile logs
S //Total no of sessions
T //Total no of music items
$M=\{m_1,....,m_{|m|}\}$;
$m_j \in M$; s= 1……S and j = 1 . . . T;
OUTPUT: C //Customer action set
ALGORITHM:
for each record from mobile logs
for each user from record
for (s = 1 to S)
for(j = 1 to M)
$C= C_{sj}$ **…..(1)**
Update rating whereas $C_{sj} = r$, ratings are 0,1,2,3

IV. Conclusion and Future Work

The rows in the customer action set increase according to the frequency of the customer's visits to the site. Assume that there is a target customer (Customer A) and that a customer action set is created for the customer as follows in Table 1:
Example 1:

Phase 2: Ordinal scale-based customer profile creation

The customer profile describes a customer's product interests or preferences. In this phase, the customer profile, which represents its components on an ordinal scale, is created.

Step 2.1. Preference intensity matrix creation

A customer action set for each customer has L rows corresponding to the number of rows in the session file. Individual rows of customer action sets contain a part of the preference information, and all must be aggregated to make a unified customer profile over all music items. To incorporate preferences, including the degrees, the term preference intensity matrix is used [5]. Shows in Table2:

Preference Intensity Matrix Creation algorithm
INPUT: C // Customer Action Set
OUTPUT: P^l //Preference Intensity Matrix
ALGORITHM:
Customer action set for each customer has L rows
For (l = 1 to L)
For (i = 1 to T)
For (j = 1 to T)
$P^l = P_{ij}^l$
Where $p_{ij}^l = C_{li} - C_{lj}$ …..(2)
L = Total number of rows in C
T = Total number of music items
$\boldsymbol{p_{ij}^l}$ = Represents the relative strength/weakness of preference that a certain music item m_i has in comparison with other music items m_i in the l^{th} session.

Table 2 Preference Intensity Matrix P^1 for the 1^{st} row in Example 1

Session	Music					
	m_1	m_2	m_3	m_4	m_5	m_6
m_1	0	1	1	3	3	2
m_2	-1	0	0	2	2	1
m_3	-1	0	0	2	2	1
m_4	-3	-2	-2	0	0	-1
m_5	-3	-2	-2	0	0	-1
m_6	-2	-1	-1	1	1	0

Step 2.2. Optimal preference intensity matrix creation

After Step 1, a certain customer's L individual preference information about music is identified with degrees by L corresponding preference intensity matrices $\{P^l\}_{l=1}^{L}$. In this step, preference intensity matrices $\{P^l\}_{l=1}^{L}$ for a specific customer are put together. And then, an optimal preference intensity matrix $\hat{X}$, which implies comprehensive preference information on that customer's music preferences, is generated. Then, $\hat{X}$ can be defined as follows:

Example 3: According to 3, the target Customer A's optimal preference intensity matrix is as follows in Table 3:

Optimal preference intensity matrix creation algorithm
INPUT: $\{P^l\}_{l=1}^{L}$ // Preference Intensity Matrix
OUTPUT: $\hat{X}$ // Optimal Preference Intensity Matrix
ALGORITHM:
The optimal solution is,
for each preference intensity matrices p_1 to p_L
X^*_{ij} = median $\{P^l_{ij}$, l= 1,…..L$\}$, i < j. ….(3)

Table 3
Optimal Preference Intensity Matrix for Customer A

	m_1	m_2	m_3	m_4	m_5	m_6
m_1	-	1	1	3	2	2
m_2	-	-	0	2	0	1
m_3	-	-	-	2	1	1
m_4	-	-	-	-	-1	-1
m_5	-	-	-	-	-	0
m_6	-	-	-	-	-	-

Step 2-3. Ordinal scale-based customer profile creation

Finally, the ordinal scale-based customer profile for our recommender systems is created in this step. This requires a series of transformations that begin from the optimal preference intensity matrix. As mentioned above, value of pij in the optimal preference intensity matrix contains the relative strength/weakness of an individual customer's preference between mi and mj. Therefore, by reversely iterating Step 1 to the optimal preference intensity matrix, we arrive at the sorted sequence Ob, which is denoted as Ob = m1 > m2 > m3 > … > m|Mb|. This indicates that customer b prefers music m1 to m2 and prefers m2 to m3, and so on. The order that underlies this sorted sequence must be declared in terms of ordinal numbers. The rank r(Ob,mj) refers to the cardinal number indicating the position of the music item mj in Ob. For example, Ob = m1 > m3> m2, r(Ob,m2) is 3. The sequence of rankings Eb = {r(Ob,mj)|mj $\in$ Mb} denotes a set of ranking of music for customer b. Eb is a component of the ordinal scale-based customer profile representing customer b's specific preferences. Therefore a complete customer profile U = {E1,..,EY} is created by iterating these three steps repeatedly for all Y customers.

Example 4: From the optimal preference intensity matrix (Table 3), Customer A's OA = m1 > m2, m3 > m5, m6 > m4 and EA={ 1 , 2 , 2 , 4 , 3 , 3 } are created. Suppose that there are five more Customers B, C, D, E, and F and that all the steps in Phase 2 are iterated for these customers to make complete customer profiles. The ordinal scale-based customer profile, U, is thus created as follows in Table 4:

Table 4
Sample of the customer profile

Customer	Music										
	m_1	m_2	m_3	m_4	m_5	m_6	m_7	m_8	m_9	m_{10}	m_{11}
A	1	2	2	4	3	3					
B	2	2	1	3	2	3	1	2		3	1
C	1	3	2		3	1	3	1	2		
D	2	1		4	3	3		1	4		3
E	1	2	3	3	3	2	3	3	4	4	4
F		2		3	2	4		1	4	1	2

PHASE 3: NEIGHBORHOOD FORMATION AND RECOMMENDATION GENERATION

Given the customer profile, U, the CF-based recommendation procedure for a target customer is performed in this phase.

STEP 3.1. NEIGHBORHOOD FORMATION

This step computes the similarity between customers and forms a neighborhood between a target customer and a group of like-minded customers [10]. The similarities between the target customer a and other customers b are as follows:

$$R_{ab} = \frac{\sum_{m_j \in M_{ab}} (r(O_a, m_j) - \overline{r_a})\,(r(O_b, m_j) - \overline{r_b})}{\sqrt{\sum_{m_j \in M_{ab}} (r(O_a, m_j) - \overline{r_a})2}\sqrt{\sum_{m_j \in M_{ab}} (r(O_b, m_j) - \overline{r_b})2}} \quad \ldots.. (6)$$

Where $M_{ab} = M_a \cap M_b$ and $\overline{r_b} = \frac{\sum_{m_j \in M_{ab}} (r(O_b, m_j))}{|M_{ab}|}$

Using the similarity measure above, this phase determines which previous customers will be used in the recommendation for the target customer.

Example 5: Suppose that Customer A is a target customer. We can calculate R_{Ab} between Customer A and others by using the customer profile: R_{AB} = 0:63; R_{AC} = 0:30; R_{AD} = 0:81; R_{AE} = 0:70, and R_{AF} = 0:43. Assuming that n = 3 in the best-n-neighbor technique, Customers B, D, and E are then selected to comprise the neighbourhood of Customer A.

STEP 3. 2. RECOMMENDATION GENERATION

This step derives the top-N recommendation from the neighbourhood of customers. For each customer, produce a recommendation list of N music items that the target customer is most likely to download. Previously downloaded music items are excluded from the recommendation list to broaden each customer's download patterns or coverage. The preference for music j by the target customer a is estimated by the following function:

$$\hat{r}_{aj} = \frac{\sum_{b \epsilon \widetilde{M_j}} R_{ab} \times (r\,(O_b m_j) - \overline{r_b})}{\sum_{b \epsilon \widetilde{M_j}} |R_{ab}|} \quad \ldots..7$$

Where $\widetilde{M_j}$ = { b | E_b ϵ U s.t m_j ϵ M_b }

The music items are sorted in an ascending order of these estimated preferences, and the highly ranked N items are recommended.

Example 6: According to the results in Example 5, we can calculate $\hat{r}_{Ai}$ of Customer A for all of the music items:

$\hat{r}_{A1}$=1.64, $\hat{r}_{A2}$=1.58, $\hat{r}_{A3}$=2.07, $\hat{r}_{A4}$=3.34, $\hat{r}_{A5}$=2.67, $\hat{r}_{A6}$=2.63, $\hat{r}_{A7}$=2.07, $\hat{r}_{A8}$=1.92, $\hat{r}_{A9}$=3.74, $\hat{r}_{A10}$=3.55, $\hat{r}_{A11}$=2.70.

Previously downloaded items by Customer A, such as (m1; m2; m3; m5), are excluded from the recommendation list to broaden Customer A's download patterns. Thus, assuming that N is 3, we can generate a recommendation list R= (m8; m7; m6).

V. CONCLUSION

In this paper, new CF-based recommendation methodology for mobile Web music is proposed. Compared to other CF-based recommender systems, proposed system is characterized by three aspects: (1) it implicitly captures preference information by using the mWUM technique. (2) It represents customer preference on an ordinal scale.(3) In order to compromise several pieces of preference information, system applies a famous consensus model called the CK method, which has been widely used in the area of MCDM to solve ordinal consensus-making problems. As a realistic solution for music recommendation problems in the mobile Web environment, system offers the following benefits to both consumers and suppliers of mobile music. (1) Customers will be able to download content with much lower connection time on the mobile Web because they will be able to easily find the desired items; and (2) mobile content providers will be able to improve their profitability and revenues because their download conversion rate will be improved through increased customer satisfaction.

REFERENCES

1. C. C. Aggarwal, P.S. Yu, Data mining technique for personalization, Data Engineering 23 (1) (2000).
2. Y.H. Cho, J.K. Kim, Application of Web usage mining and product taxonomy to collaborative filtering in e-commerce, Expert Systems with Applications 26 (2004) 233–246.
3. C.Y. Kim, J.K. Lee, Y.H. Cho, D.H. Kim, VISCORS: a visual-content recommender for the mobile Web, IEEE Intelligent Systems 19 (3) (2004) 32–38.
4. W.D. Cook, Distance-based and ad hoc consensus models in ordinal preference ranking, European Journal of Operation Research 172 (2) (2006) 369–385.
5. W.D. Cook, M. Kress, Ordinal ranking with intensity of preference, Management Science 31 (1) (1985) 26–32.
6. B. Sarwar, G. Karypis, J.A. Konstan, J. Riedl, Application of dimensionality reduction in recommender system – a case study, in: Proceedings of the ACM WebKDD Workshop, 2000.
7. J. Cho, K. Kwon, Y. Park, Collaborative filtering using dual information sources, IEEE Intelligent Systems 22 (3) (2007) 30–38.

8. L. Chen, F. Hsu, M. Chen, Y. Hsu, Developing recommender systems with the consideration of product profitability for sellers, Information Sciences 178 (4) (2008) 1032–1048.

9. M. Claypool, A. Gokhale, T. Miranda, P. Murnikov, D. Netes, M. Sartin, Combining content-based and collaborative filters in an online newspaper, in: ACM SIGIR'99 Workshop on Recommender Systems, Berkeley, CA, 1999.

10. T. Kamishima, Nantonac collaborative filtering: recommendation based on order response, in: Proceedings of the Ninth International Conference on Knowledge Discovery and Data Mining, 2003, pp. 583– 588.