AN ENHANCED RECOMMENDATION SCHEME FOR ONLINE GROCERY SHOPPING

Yi-Jing Wu and Wei-Guang Teng
Department of Engineering Science
National Cheng Kung University
Tainan, Taiwan
wgteng@mail.ncku.edu.tw

ABSTRACT

Online grocery shopping becomes more and more popular in recent years. To facilitate the purchase process, many online stores provide a shopping recommendation system for their consumers. So far, the generic recommendation systems mainly consider preferences of a consumer based on his/her purchase histories. Nevertheless, it is noted that there is nothing to do with the right timing to purchase a product from the view point of product replenishment or economic purchasing. Hence, we develop a new recommendation scheme especially for online grocery shopping by incorporating two additional considerations, i.e., *product replenishment* and *product promotion*. We believe that such a new scheme should be able to provide a better recommendation list which fit consumer desires, needs, and budget considerations and finally boost transactions.

1. INTRODUCTION

Grocery shopping is undoubtedly one of the most frequent and necessary works of every family. However, as the life pace becomes faster and faster, people are less likely to spend time and energy on doing it. Fortunately, thanks to the vigorous development of e-commerce, people are now able to fulfill this work through online shopping. Moreover, people can use not only computers but also various types of handheld devices, e.g., PDAs, smart phones and tablets, to surf websites so as to do their shopping easily as information technology advances recently. As a result, shopping groceries online becomes more and more popular. Under such circumstance, how to make online purchasing quick and efficient becomes a vital issue in e-commence.

In view of this, most online stores provide a shopping recommendation system for the consumers to facilitate online shopping. The core of such systems is a personalized recommendation algorithm. This algorithm models consumer shopping behaviors and recommend items to the consumers while doing on-line purchasing. Since there is no explicit product rating available for grocery shopping, the system has to estimate consumers' preferences from their purchased histories. One of the major techniques used to develop a recommendation algorithm is collaborative filtering (CF). Nevertheless, the problem of sparsity due to too few user ratings may make the formation of neighborhood inaccurate and thereby results in poor recommendations.

In this work, a methodology based on random walk and bipartite networks [9] is utilized and adapted to alleviate the problem of sparsity. To the best of our knowledge, conventional models cannot be directly applied in our target application of grocery shopping. Some other factors should also be considered. For instance, most products in grocery stores are daily necessities which are consumable and are purchased periodically. Therefore, product replenishment must also be considered in addition to user preferences. Besides, in real life, product price strongly affects consumers' purchase willingness, especially for budget consumers. Therefore, product promotion plays a very important role for decision-making. In other words, one may easily understand that a recommendation system without considering product promotion is not practical for the application of grocery shopping.

Regarding the issues mentioned above, we develop a novel grocery shopping recommendation scheme by incorporating two additional considerations, i.e., *product replenishment* and *product promotion*, with the generic recommendation system which considers about product similarities and individual interest only. In addition, to enhance the estimation of a consumer's individual interest in this work, we divide consumer online purchasing behaviors into three steps of "viewing the product information," "adding product to shopping basket" and "purchasing product." Such a new scheme should be able to provide a more appropriate recommendation list which fit consumer desires, needs, and budget considerations.

The rest of this paper is organized as follows. Preliminaries and related works are reviewed in Section 2. Our proposed methods and scheme used to estimate consumers' preference are introduced in Section 3. The prototyping of our proposed recommendation system for on-line grocery shopping is explored in Section 4. Finally, this paper concludes with Section 5.

978-1-61284-842-6/11/$26.00 ©2011 IEEE

2. PRELIMINARIES

Common recommendation techniques are explored in Section 2.1. Moreover, previous works on supporting consumers in their grocery shopping process are reviewed in Section 2.2.

2.1 Common Recommendation Techniques

Recently, the development of recommendation systems has attracted significant research interests, especially in the field of e-commerce. In recommendation systems, user preferences are estimated so as to provide a recommended items-list for the users. Specifically, the recommendations can be made based on either user purchase logs or inferences from other users having similar preferences. Examples applications include recommending books, CDs, and gifts at Amazon.com [10].

Item-based CF [5][11] is one of the most popular techniques in recommendation systems and has shown better performance in comparison with user-based CF [8][12]. Item-based CF performs the similarity calculation in item space so as to reduce the dimensionality of large-scale datasets [7]. This is because the number of consumers is typically much larger than that of items in many practical applications. However, the problem of sparsity still remains in item-based CF. In order to solve this problem, many previous techniques focus on looking for more implicit indicators [6] of consumer preferences, such as the purchasing behavior [1].

Currently, many recommendation techniques are based on user ratings which are explicitly specified by the users to represent the degree of their preferences. The ratings are usually represented on a discrete numerical scale ranging from the lowest (most disliked) to the highest (most favored) value. However, for our target application of grocery shopping, it is quite difficult to get user ratings since the items are relatively cheaper and more common. Therefore, previous recommendation techniques based on user ratings cannot be directly applied in this work.

2.2 Recommendation Techniques for Grocery Shopping

Some previous works utilize association rules and clustering analysis to solve the problem that there is no explicit ratings in grocery shopping. Namely, association rules are used to whether some products are likely to be purchased together [4]. Thus, associated products can be recommended when the corresponding product is to be purchased. On the other hand, clustering analysis is to group consumers with similar purchasing histories since they may have the same preferences. Therefore, favorite products can be recommended to other members of the same group [3].

Some other prior works model purchasing histories as a bipartite network [9]. As shown in Figure 1, there are two types of nodes, i.e., consumers and products, in such a bipartite network Edges only lie between nodes of different types. Moreover, the strength of an edge $E(C_i, P_j)$ in this network may be used to represent the times of consumer C_i buying product P_j.

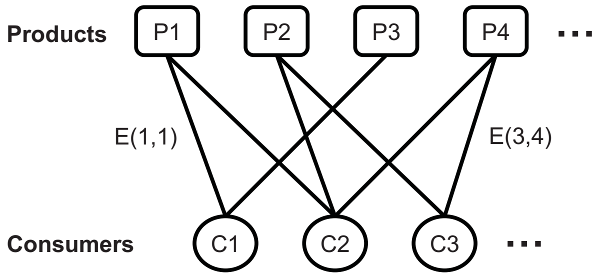

Figure 1. A bipartite network showing the consumer-product relationships.

The relevance or similarity between two nodes of the same type, i.e., two products or two consumers, can be calculated using a random walk approach [13]. This works even if two products have never been purchased together [9], thus alleviating the problem of sparsity. In view of this, the corresponding techniques are adapted in this work to estimate product similarities and individual interests.

3. PROPOSED SCHEME FOR ONLINE GROCERY SHOPPING

General concepts of our proposed approach are introduced in Section 3.1. Corresponding technical details are illustrated in Section 3.2. Moreover, the overall recommendation scheme for online grocery shopping is presented in Section 3.3.

3.1 General Concepts of the Proposed Approach

For grocery shopping, consumer preferences could be formed by combining three aspects, i.e., *individual interest*, *product replenishment*, and *product promotion*. Once the preferences can be precisely estimated, a corresponding recommendation technique is of high potential to be utilized in practical applications.

Firstly, a bipartite network is constructed based on purchasing histories of all users to estimate the individual interest. Note that consumers are likely to accept product recommendations that are similar to what they have bought before. Also, it is observed that one is willing to accept recommendations from consumers of similar tastes.

Thus, by introducing the random walk approach, how much a specific consumer likes a product can be generally estimated.

Secondly, most daily necessities are consumable and are targets of grocery shopping. Therefore, consumers may buy the same product repeatedly. This is regarded as "product replenishment." Note that consumable products normally are with a constant consumption rate. People then have to purchase them periodically. When something is going to be exhausted, the purchasing intent of a consumer becomes firm. In this work, we thus propose a statistical model to estimate this factor of product replenishment.

Finally, a most common strategy to increase the sales of a product is promotion. People always like to do their purchases at a reduced price. Therefore, to consider the effect of product promotion in the recommendation system is necessary. Specifically, we model the degree of product promotion as well as the customer sensitivity of money saving to estimate the willingness for a consumer to buy a specific product.

3.2 Estimation of Consumer Preferences

As mentioned in previous sections, product similarities can be calculated based on the weights of all edges in a bipartite network. Suppose a matrix of product similarities $\boldsymbol{P}$ is obtained, the random walk approach is then utilized. Equation (1) is iterated to obtain the ranking scores for all the products based on the current items in the consumer's basket [9].

$$\boldsymbol{R_{basket}} = d \cdot \boldsymbol{P} \cdot \boldsymbol{R_{basket}} + (1-d) \cdot \boldsymbol{U_{basket}} \quad (1)$$

In Equation (1), $\boldsymbol{R_{basket}}$ is the basket-based ranking score vector used for ranking all the products, and the *i-th* entry of the initial vector $\boldsymbol{U_{basket}} = 1/m$, if the *i-th* product is in the basket and 0 otherwise, $d \subset (0, 1)$ is a damping factor, and m is the number of products in the current basket.

Since the ranking score vector $\boldsymbol{R_{basket}}$ coming out from iterations of Equation (1) is based on the products in the basket, therefore, two consumers with the same content in their baskets have same items-list to be recommended. This situation is unreasonable from the view point of consumer personalization. In order to provide personalized recommendations to a consumer, the vector of individual interest is assigned with different weights as compared to $\boldsymbol{R_{basket}}$. Namely, the individual interest vector $\boldsymbol{I}$ of a specific consumer is calculated by Equation (2).

$$I(c_k, p_i) = \frac{1}{3}\left(\frac{N_v(c_k, p_i)}{\sum N_v(c_k, \cdot)} + \frac{N_b(c_k, p_i)}{\sum N_b(c_k, \cdot)} + \frac{N_p(c_k, p_i)}{\sum N_p(c_k, \cdot)}\right) \quad (2)$$

To formulate Equation (2), the on-line purchasing process of consumer k is divided into three stages of "viewing the product information," "adding product to shopping basket" and "purchasing product" [4]. N_v is the number of consumer k to viewing the product i. N_b is the number of consumer k to adding the product i to shopping basket. N_p is the number of consumer k to purchase the product i. Therefore, $I(c_k, p_i)$ is the interest vector of consumer k with product i. In conclusion, consumer preference CP_1 is obtained by summing up $\boldsymbol{R_{basket}}$ ranking scores vector and individual interest vector $\boldsymbol{I}$.

The probability density distribution for a consumer to replenish a consumable product as function of the time interval t between two consecutive purchases to the same category of products can be a gamma distribution [2][14]. The probability density function of a gamma distribution can be formulated as Equation (3).

$$f(X = t; \alpha, \beta) = \frac{1}{\beta^{\alpha}\Gamma(\alpha)} x^{\alpha-1} e^{-x/\beta}, \quad x \geq 0 \quad (3)$$

In Equation (3), α is the shape parameter and β is the scale parameter. In this study, we calculate the averaged purchase time interval of a particular product, parameters of α and β based on consumer's purchased histories. Then we input these values to the developed model of gamma distribution to do computation and get the probability of purchase to the same product in this visit. Here, the probability of purchase is the consumer preference CP_2. This probability is based on the condition that this particular product has not been purchased for t days so far. In practice, a consumer may purchase different amount of a product at each purchase. Therefore, it is very important to consider the average purchase amount and the average purchase interval.

Note that product price as well as money saving strongly affect the purchase willingness of a consumer. But these factors are not considered in conventional shopping recommendation systems. With the absence of these factors, the shopping recommendation system could not provide precise recommendations to the consumers, especially to the budget consumers. Hence, in this work, we model the effect of these two factors, price and money saving, to be the third part of consumer preference CP_3.

The willingness of purchasing a product on promotion is relevant with the discount rate *(D)* of a product and the consumer's sensitivity *(S)* of money saving. As some products are often on sale and some others are seldom discounted, both the actual money saving and the average price should be taken into consideration when calculating the discount rate *(D)* of a product. For example, a product with an original price of \$100, but it is quite often to be sold with a discount rate of 10%. Moreover, the average price of this product during a whole year is \$95. For this case, if the price of this product is \$90 now, then its discount rate is (100-90) / 95 = 10.5%. On the other hand, products which are seldom discounted could be of higher priority in the recommended item-list.

Besides, different consumers may be with different sensitivities to the matter of money saving. Intuitively, the sensitivity to a \$40 saving could be taken as high to a consumer with a weekly consumption of \$200, but low to a consumer with a weekly consumption of \$2000. Therefore, Sensitivity (*S*) can be set as the value of money saving divided by consumer's averaged weekly consumption. Consumer preference CP_3 is calculated by summing up above two psychological effects together with different weight assigned to these two effects.

Normally, before the consumer selects a product and put it into the shopping basket, he/she may consider the preferences of this product as well as the necessity of replenishment and degree of promotion. To effectively simulate this thinking process, we select certain numbers of products based on the ranks of the first consumer preference CP_1. Then, we re-arrange these products' sequence based on consumer preferences CP_1, CP_2, and CP_3, and assign a new ranking value to each selected product (rank-*a*, rank-*b*, rank-*c*). After that, we sum up three ranking values (rank-*a*, rank-*b*, rank-*c*) of each product together to get the final ranking values to generate the final recommendation list.

3.3 Our Recommendation Scheme

As shown in Figure 2, the proposed scheme can be divided into four parts, including *information manager, database, analyzer*, and *recommender*. The functionalities of each part is illustrated as follows.

(1) *Information manager*: This part is to build consumer profiles such as gender, age, occupation, etc. On the other hand, when the consumer logs in to purchase products, the information manager begins to record the consumer behavior logs including the browsing time and clicks.

(2) *Database*: This part is to store all on-line purchasing behavior and purchase histories of a consumer. Also, promotion plans of all the products are stored.

(3) *Analyzer*: This part is to analyze database contents for obtaining product similarities, individual interests, replenishment time intervals, promotion degree so as to estimate consumer preferences. We use these consumer preferences to comprehensively evaluate the needs of a consumer.

(4) *Recommender:* This part is to produce a recommendation list for a specific consumer according to the analysis result. The recommendation list shows the products which are most likely to be purchased. In general, this list includes not only replenishment goods but also products on promotion.

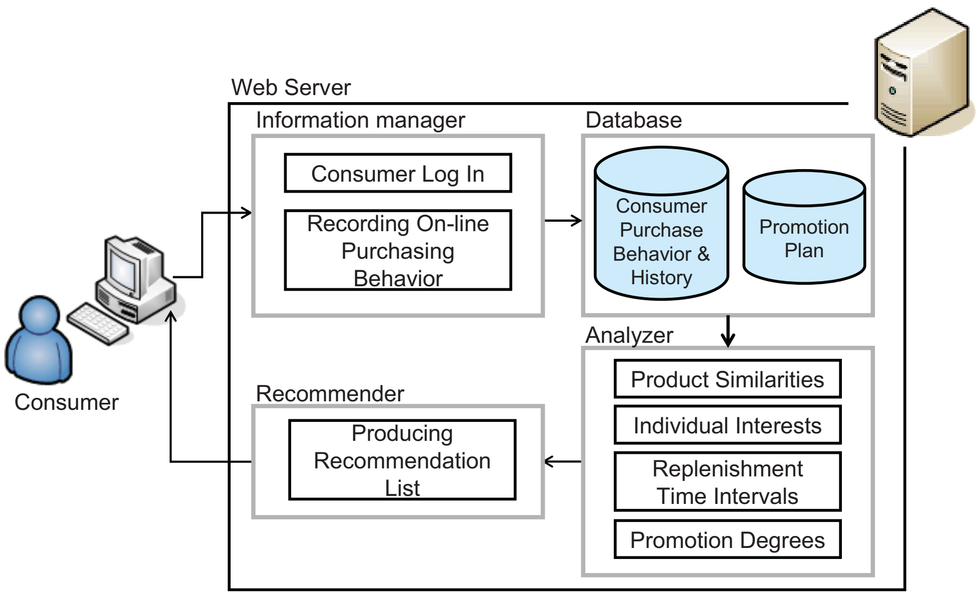

Figure 2. Proposed recommendation scheme for on-line grocery shopping.

4. PROTOTYPING OF PROPOSED RECOMMENDATION SCHEME

A prototype of the proposed scheme is introduced in Section 4.1. In addition, a case study is provided in Section 4.2.

4.1 Prototyping of an On-line Grocery Recommendation System

In this work, we develop a prototype of the proposed scheme to facilitate the empirical studies. Note that product information are obtained from real online stores. Consumers can thus browse the up-to-date information of all real products. The user interface is as shown in Figure 3. A consumer can search for interested product or select a product category to begin his/her browsing and online purchasing. A significant difference of our scheme from generic online stores is that our website considers not only the individual interest, but also the product replenishment as well as money saving for customers.

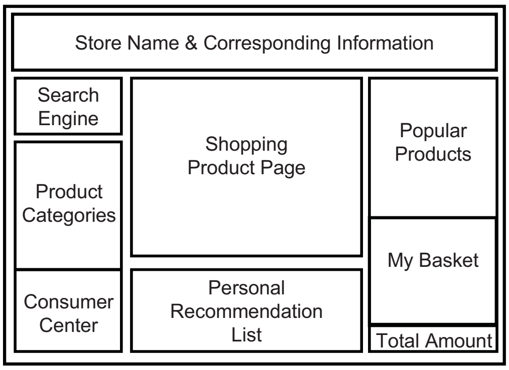

Figure 3. The user interface of proposed on-line grocery recommendation system.

When the consumer starts a shopping session, the shopping basket is empty. Consumer preferences CP_1 have no value. Therefore, our scheme sums CP_2 and CP_3 up together to get the overall preference ranks and then selects top five products to make recommendation. Subsequently, the consumer put the milk into shopping basket. The proposed scheme computes CP_1 based on the current basket and the purchase histories. Thus it can be seen that milk and toast are most frequently purchased together. However, the consumer purchased milk even two days before. The system compute the rank of CP_2 for toast is low. Besides, the rank of CP_3 for toast is also low. Consequently, toast is not recommended. This example is further illustrated in Figure 4.

Current Basket

Rank	CP_1	CP_2	CP_3
1	Toast	Yogurt	Bread
2	Cereal	Cereal	Soap
3	Yogurt	Rice	Cheese
4	Bread	Cheese	Jam
5	Jam	Toilet paper	Yogurt
6	Cheese	Bread	Rice
7	Coke	Jam	Toilet paper
8	Soap	Toast	Cereal
9	Toilet paper	Coke	Toast
10	Rice	Soap	Coke

Recommendation List

Rank	CP_f
1	Yogurt
1	Bread
3	Cereal
4	Cheese
5	Jam

Figure 4. An example for providing recommendation list by combining three consumer preferences

4.2 A Case Study

A common event in our daily life is taken as an example to illustrate the feasibility of our scheme. A consumer, Mrs. Lin, is a busy housewife and used to purchase daily necessities from online stores. Assume that Mrs. Lin visits a conventional online grocery store on 2011-04-25 that is flooded with product information and promotion plans. Also, only popular products or products which are most frequently purchased by her are recommended. Suppose all these recommended products are as listed in Table 1. There is a potential problem in this case. We can observe from Table 1 that she just bought milk and toilet paper three days ago. Therefore, there is low probability that Mrs. Lin runs out of these two products and needs to purchase them again at this time. In other words, to recommend milk and toilet paper to Mrs. Lin for purchasing at this time is inappropriate. Besides, this store may already launch promotional plans for some products at this moment. But there is no way to notify Mrs. Lin.

Once our scheme is introduced in this scenario, the problem as mentioned above can be eased. Specifically, the consumer preference CP_1 is firstly estimated according to the purchasing logs of all consumers. Thus, products with the highest values of CP_1 are listed. Suppose that these products are identical to the ones as shown in Table 1. Then, further calculation based on the factors of product replenishment (CP_2) and product promotion (CP_3) are performed. After all these products are ranked according to the value of CP_1, CP_2 *and* CP_3, respectively, the overall ranking score can be obtained by summing up these three individual rankings. Finally, as shown in the last two columns of Table 1, coke and cereal are the top two products to be recommended. On the other hand, milk and toilet paper are the top two products which are unlikely to be purchased among these six products.

Table 1. The information of Mrs. Lin's most frequently purchased products

Product Name	*Last Purchase Date*	*Mean Time Interval (day)*	*Original Price (NTD)*	*Current Price (NTD)*	*Average Price (NTD)*	*Rank of CP_1*	*Rank of CP_2*	*Rank of CP_3*	*Score*	*Rank of CP_f*
Milk	2011-04-22	7	60	60	55	1	4	6	11	3
Toilet paper	2011-04-22	20	120	115	108	2	6	5	13	6
Diaper	2011-04-17	25	220	180	210	3	5	4	12	4
Coke	2011-04-17	5	57	40	55	4	1	1	6	1
Cereal	2011-04-12	15	90	65	85	5	2	2	9	2
Rice	2011-04-08	30	130	100	117	6	3	3	12	4

5. CONCLUSIONS

In this work, we have proposed to develop a recommendation technique for online grocery shopping. Specifically, three different factors, i.e., individual interests, product replenishment, and product promotion, have been considered in the proposed scheme. In addition, on-line purchasing behavior has been separated into three stages, i.e., "viewing the product information," "adding product to shopping basket" and "purchasing product." Therefore, a more appropriate recommendation list can be generated to fit consumer desires, needs, and budget considerations.

6. ACKNOWLEDGMENT

The authors are supported in part by the National Science Council, Taiwan, R.O.C. (NSC98-2221-E-006-164-MY2)

7. REFERENCES

[1] S. Ahmed, P. Ko, J.-W. Kim, Y.-K. Kim, and S. Kang, "An Enhanced Recommendation Technique for Personalized E-Commerce Portal," *Proceedings of the Second International Symposium on Intelligent Information Technology Application*, pages 196-200, December 2008.

[2] G. M. Allenby, R. Leone, and L. Jen, "A Dynamic Model of Purchase Timing with Application to Direct Marketing," *Journal of the American Statistical Association*, pages 365-374, June 1999.

[3] Y. B. Cho, Y. H. Cho, and S. H. Kim "Mining Changes in Customer Buying Behavior for Collaborative Recommendations," *Expert Systems with Applications*, 28(2):359-369, February 2005.

[4] Y. H. Cho and J. K. Kim, "Application of Web Usage Mining and Product Taxonomy to Collaborative Recommendations in E-commerce," *Expert Systems with Applications*, 26(2):233–246, February 2004.

[5] M. Deshpande and G. Karypis, "Item-based Top-N Recommendation Algorithms," *ACM Transactions of Information System*, 22(1):143–177, January 2004.

[6] S. Gadanho and N. Lhuillier, "Addressing Uncertainty in Implicit Preferences," *Proceedings of the 2007 ACM Conference on Recommender Systems*, pages 97-104, November 2007.

[7] C.-N. Hsu, H.-H. Chung, and H.-S. Huang, "Mining Skewed and Sparse Transaction data for personalized shopping recommendation," *Machine Learning*, 57(1-2):35-59, October 2004.

[8] J. A. Konstan, B. N. Miller, D. Maltz, J. L. Herlocker, L. R. Gordon, and J. Riedl, "Grouplens: Applying Collaborative Filtering to Usenet News," *Communications of the ACM*, 40(3):77–87, March 1997.

[9] M. Li, B. M. Dias, I. Jarman, W. El-Deredy, and P. J. Lisboa, "Grocery Shopping Recommendations Based on Basket-Sensitive Random Walk," *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 1215-1224, July 2009.

[10] G. Linden, B. Smith, and J. York, "Amazon.com Recommendations: Item-to-Item Collaborative Filtering," *IEEE Internet Computing*, 7(1):76-80, February 2003.

[11] B. Sarwar, G. Karypis, J. Konstan, and J. Reidl, "Item-based Collaborative Filtering Recommendation Algorithms," *Proceedings of the 10th International Conference on World Wide Web*, pages 285–295, May 2001.

[12] U. Shardanand and P. Maes, "Social Information Filtering: Algorithms for Automating Word of Mouth," *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 210–217, May 1995.

[13] J. Sun, H. Qu, D. Chakrabarti, and C. Faloutsos, "Relevance Search and Anomaly Detection in Bipartite Graphs," *SIGKDD Explorations Newsletter*, 7(2):48-55, December 2005.

[14] R. Venkatesan and V. Kumar, "A Customer Lifetime Value Framework for Customer Selection and Resource Allocation Strategy," *Journal of Marketing*, 68(4):106-125, October 2004.