Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 91 (2016) 944 – 952

Procedia Computer Science

Information Technology and Quantitative Management (ITQM 2016)

Implementation of a Recommendation System using Association Rules and Collaborative Filtering

"JinHyun Jooa, SangWon Bangb, GeunDuk Parka" *

[a]Hoseo University, 20, Hoseo-ro, 79 beon-gil, Baebang-eup, Asan-si 31499, Korea
[b] Hunter Technology, 1408, 1st Woorim e-Biz Center 28, Digital-ro 33-gil, Guro-gu, Seoul, Korea

Abstract

In this study, a recommendation system was designed and implemented which analyzes using patterns and personal propensities of customers by using association rule analysis and collaborative filtering for collected customer data on visiting customer companies with NFC (Near Field Communication). The recommendation algorithm used in the proposed system used the data analysis results and the distance data from GPS (Global Positioning System) to recommend local businesses that people are highly likely to visit.

Keywords: Association Rule; Collaborative Filtering; Recommendation Systems; NFC

1. Introduction

Ericsson, a Swedish multinational telecommunications corporation, predicted in their Ericsson Mobility report that by 2019 the number of smartphone subscribers will reach 5.6 billion. If this forecast is accurate, most of the world's people will be using smartphones and smart devices by then. [1] Eric Schmidt, the CEO (Chief Executive Office) of Google, said that today, the amount of data generated in the world every two days was equivalent to the amount of data from the start of human civilization to the year 2003. [2]

It is extremely difficult to find the data which a user needs in this sea of data that are generated according to smartphone usage. Therefore, it is necessary to sort information needed by the user among much data. For such, researchers had conducted many studies with particular interests in personalization and recommendation systems. Personalization refers to providing users with products or services according to their characteristics and tastes without explicitly asking them. [3]

* Corresponding author. Tel.: fax: +82-41-548-9667.
E-mail address:gdpark@hoseo.edu

doi:10.1016/j.procs.2016.07.115

Collaborative filtering is one of the most commonly-used methods in the area of recommendation systems. [4] In collaborative filtering, the customers' preference data for products are collected and a customer is recommended products that they are likely to purchase based on their preferences.

There are two methods of collecting data for collaborative filtering, which are explicit and implicit. The implicit method obtains data based on user behavior or history although its accuracy is slightly lower than the explicit method where the customer is asked directly, but has an advantage that could reflect their changing tendencies.

In this study, user data are implicitly collected via NFC (Near Field Communication). The advantage of NFC is that most smartphones have a built-in NFC functionality, and only tags are needed to use it. The use of NFC allows to improve inconvenience of loosing or keeping paper coupons or RFID card coupons.

This study collects customer information implicitly and predicts seller with high possibility of customer visit through association rule analysis and collaborative filtering by using NFC. Also, the user's preferred business that is the closest to their current location is recommended based on GPS data, and a personalized recommendation system that maximizes user convenience is implemented, where the user does not have to try to find the closest business manually. In this study, the overall process of the association rule analysis and the collaborative filtering are examined for the recommendation system. Furthermore, a test is done to see whether this implemented system is able to find the appropriate businesses for the user.

2. Related Research

2.1. Recommendation system

The recommendation system is a service that analyzes customer data, including the user's purchase data, in order to recommend them the most suitable products or services.[3][5] Most recommendation systems have the process flow. By analyzing the gathered customer data, similarities between users are found, and finally items that the user is likely to purchase are recommended. The system can collect various types of data, such as usage patterns, by analyzing their demographic information, purchase patterns, and click streams as the users access the Internet on their smartphones; these data can be collected by requiring membership, where the user has to sign up. Such recommendation systems are being actively studied in various areas such as online transactions and location-based services, and they are used in online e-commerce sites such as Amazon and Netflix, as well as in smartphone apps.

2.2. Collaborative filtering

Collaborative filtering is a way to automatically predict customers' interests based on the taste information gathered from many customers. The premise of collaborative filtering is that the current behavior of customers will be maintained in the future. [6] That is, in collaborative filtering, a small group of similar customers is found, and a list of recommended items is created, comprised of items that the user is most likely to select. There are two major approaches to collaborative filtering, user-based and item-based. In the user-based approach, the list of recommended items is created based on the customers, while in the item-based approach the list is created based on the products. In the user-based approach, similarity between users is used to predict the levels of product preference. In the item-based approach, similarity between different products is calculated based on the products that users have reviewed in order to predict the levels of product preference. With the item-based collaborative filtering, the algorithm is complicated and involves a lot of calculations because the similarity is estimated for each item, so that the accuracy of the algorithm is poor when users have differing preferences. The user-based collaborative filtering, on the other hand, has fast processing speeds and is easy to implement.

2.3. Association rules

Data mining is a method of finding hidden relationships, patterns and rules in massive amounts of data in order to extract useful information from the data, in the readily-understandable form of new rules, tendencies, and patterns. The information so extracted can be useful for business marketing strategies or when it comes to customers making decisions. Furthermore, the type of data mining method most actively being researched is the one using association rules. Association rules specify the rules about how one event is related with another. It is also a type of clustering that classifies data by relevance. [7] In its formal definition, database D is defined as a set of transactions, $D = \{T_1, T_2, \dots, T_n\}$, and transaction T is defined as a set of items that make up the transaction, $T = \{ i_1, i_2, \dots, i_n \}$. Duplicate transactions are not allowed, and transactions are assumed to be sorted. Also, an association rule is expressed in the form $R: A \rightarrow B$, which indicates that when Event A occurs Event B may occur. In the association rule $R: A \rightarrow B$, X is called the antecedent and Y the consequent of the rule, with $X \cap Y = \emptyset$. [7][8][9] Two scales are used to assess the relations in association rules, support and confidence. In most of the association rule discovery algorithms, support is the ratio of the rule R to the total number of transactions, while the confidence refers to the strength of the association, and it is the ratio of transactions that include the consequent to the total number of transactions for the event. For example, if Event X occurred in the transaction, it is the probability of Event Y also occurring.

2.4. Clustering

Cluster analysis or clustering involves forming clusters with objects. A cluster is a group of similar data objects. The objects within a cluster are more similar to one another than objects in a different cluster.

If only collaborative filtering is used to make recommendations, it would take a lot of time because large amounts of data would have to be compared to determine the nearest neighbor. To address this, clusters can be formed beforehand before the data are compared to reduce the relative size of compared data, which will eliminate the noise and the system will perform well in coming up with a list of recommended items.

2.5. Association rule based K-means algorithm

The K-means algorithm repeats n d-dimension data i times and groups them in k clusters, with complexity. Typically the values of k and i are much smaller than n, so this algorithm is highly efficient for processing large amounts of data. As for the disadvantages of K-means algorithm, it needs as an input the value for k -- the number of clusters -- and the initial weights are set arbitrarily, so that it may find the local optimum. In the association rules-based K-means algorithm, a value observed from association rules is inputted for k, and weights, which are set arbitrarily in the original K-means algorithm, are instead set with the values obtained from association rules.

3. Implementation of a Mobile Coupon Recommendation System

3.1. System Architecture

Fig. 1 shows the overall architecture of the mobile coupon recommendation system which uses NFC. The key components of the recommendation system proposed in this study are the recommendation program and the recommendation server. It is assumed that the data collection component which uses NFC and the reader has already been implemented.

The mobile coupon recommendation system is a smartphone app where the user requests for a recommendation and then receives it. Typically this kind of an app allows the user to view their purchase history and collect mileage besides the recommendation feature, but this study implemented only the recommendation feature. The recommendation server is made up of the TCP/IP-based communication manager, which receives coupon recommendation requests sent by the mobile coupon recommendation app, the database manager which processes database requests, and the recommendation manager which executes recommendation algorithms.

The communication manager is responsible for communicating with the user's smartphone. Based on TCP/IP, it processes the recommendation request packets sent from the user's smartphone. The recommendation server gets a recommendation list and sends it to the user's smartphone.

The database manager does database input/output processing, and it handles data such as the customer data and purchase history. It is invoked by the recommendation manager. It uses JDBC (Java Database Connectivity) to communicate with the database. The recommendation manager deals with the recommendation algorithm and it consists of two different major operations: usage pattern analysis using the association rule based K-means algorithm, and item recommendation for similar customers using collaborative filtering.

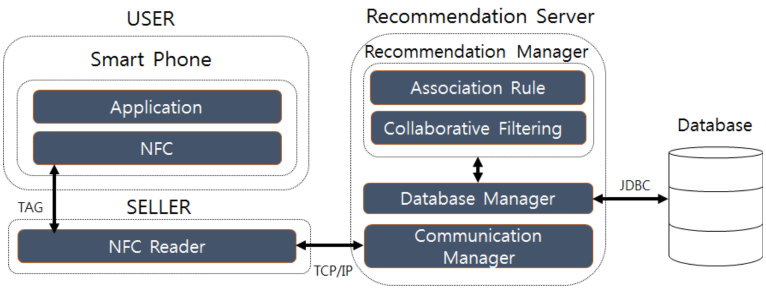

Fig. 1 System Architecture

3.2. Design and implementation of the recommendation system

The recommendation system is a Java-based server program which processes the recommendation requests received from the mobile coupon recommendation program. It is broken into two major processes.

The first process involves identifying customers' purchase patterns using an association rule based K-means algorithm in order to predict product categorizes that the customer is most likely to purchase. For example, a customer's purchase pattern may be "Customer A drinks coffee after lunch." The approval rating for the association rule will increase when there are many instances of this purchase pattern in Customer A's transactions, and the likelihood of getting selected will increase.

The second process involves using collaborative filtering to recommend items that the customers do not use, among highly similar items. The recommendation system of this study recommends items that the customers have not used because the premise is that the they can make the judgement for themselves when it comes to items that they have used, without help from the system. Therefore, collaborative filtering is used with respect to the business types found from the association analysis to make up the recommendation list for the user, containing items that are relevant for the user. The association analysis algorithm finds the purchase patterns in the customer usage data to predict their next behavior.

In this study, the Apriori algorithm, one of the algorithms for analyzing association rules, is implemented and used. The Apriori algorithm is a relatively simple and easy-to-implement algorithm. Fig. 2 shows how it is implemented. [10] Fig. 3 shows the ER (Entity-Relationship) diagram of the database which was implemented for the system. In Fig. 3, TB_USER_VISIT_INFO is a table that stores the information about customers' use of businesses. The records of this table are also tag records, and a single record is generated per use of the business. From this study, there were different transactions for breakfast, lunch, and dinner. It was assumed that customer purchases were made based on the different types of meals. Because the usage pattern differs for each type of meal, the transactions stored in the database were distinguished by breakfast, lunch and dinner. As an example of transaction preprocessing, suppose R = {A, B, D, E, C, F, G} are records of TB_USER_VISIT_INFO recorded during the course of day. Given {A}, {B}, and {D} were during breakfast, {E} and {C} during lunch, and {F} and {G} during dinner, the generated transactions are T_1= {A, B, D}, T_2= {E, C}, and T_3= {F, G}.

$$
\begin{array}{l}
\mathrm{Apriori}(T, \epsilon) \\
\quad L_1 \leftarrow \{\mathrm{large\ 1-itemsets}\} \\
\quad k \leftarrow 2 \\
\qquad \mathbf{while}\ L_{k-1} \neq \ \mathit{emptyset} \\
\qquad\quad C_k \leftarrow \{a \cup \{b\} \mid a \in L_{k-1} \wedge b \in \bigcup L_{k-1} \wedge b \notin a\} \\
\qquad\quad \mathbf{for}\ \mathrm{transactions}\ t \in T \\
\qquad\qquad C_t \leftarrow \{c \mid c \in C_k \wedge c \subseteq t\} \\
\qquad\qquad \mathbf{for}\ \mathrm{candidates}\ c \in C_t \\
\qquad\qquad\quad \mathit{count}[c] \leftarrow \mathit{count}[c] + 1 \\
\qquad\quad L_k \leftarrow \{c \mid c \in C_k \wedge \ \mathit{count}[c] \geq \epsilon\} \\
\qquad\quad k \leftarrow k + 1 \\
\qquad \mathbf{return}\ \bigcup_k L_k
\end{array}
$$

Fig. 2 Apriori algorithm

In Fig. 2, T is the set of item transactions of a customer, L is the set of frequent items, and C is the set of candidate items. Also, the minimum support level must be set in advance. The algorithm can be broadly divided into two stages, the candidate itemset C_k creation stage, where the frequent itemset is created with the union of L_{k-1} * L_{k-1}, and the frequent itemset creation stage where items above the minimum support are gathered from the set and a frequent itemset is created. The algorithm comes to an end when the candidate itemset C_k becomes an empty set as these two stages are repeated. Also, items related to customer behavior must be found from the frequent itemset in order to recommend items.

Table 1. Frequent itemset L

Frequent item	Items
L1	{A}, {B}, {C}, {E}
L2	{A, B}, {A, C}, {A, E}, {B, C}, {B, E}, {C, E}
L3	{A, B, C}, {A, B, E}, {B, C, E}

Supposing the frequent items shown in Table 1 were found, the K-means algorithm is performed according to the given association rules to reduce the data size and improve the speed and accuracy of collaborative filtering. Collaborative filtering is a way to recommend items that customers are highly likely to use. It works by determining similar customers based on the existing customer visit information. This study used the association analysis to estimate the items of the business type that the customer selected among the business types found. This kind of collaborative filtering can be divided into three stages. First for the review matrix, it contains the review ratings from customers which are typically collected explicitly. However, this study does not collect customer data explicitly, therefore, customers cannot directly review the items and compose with the traditional collaborative filtering method. Therefore, this study is composing the review matrix for number of visits by using the history of customers' visit to the business to create the evaluation matrix.

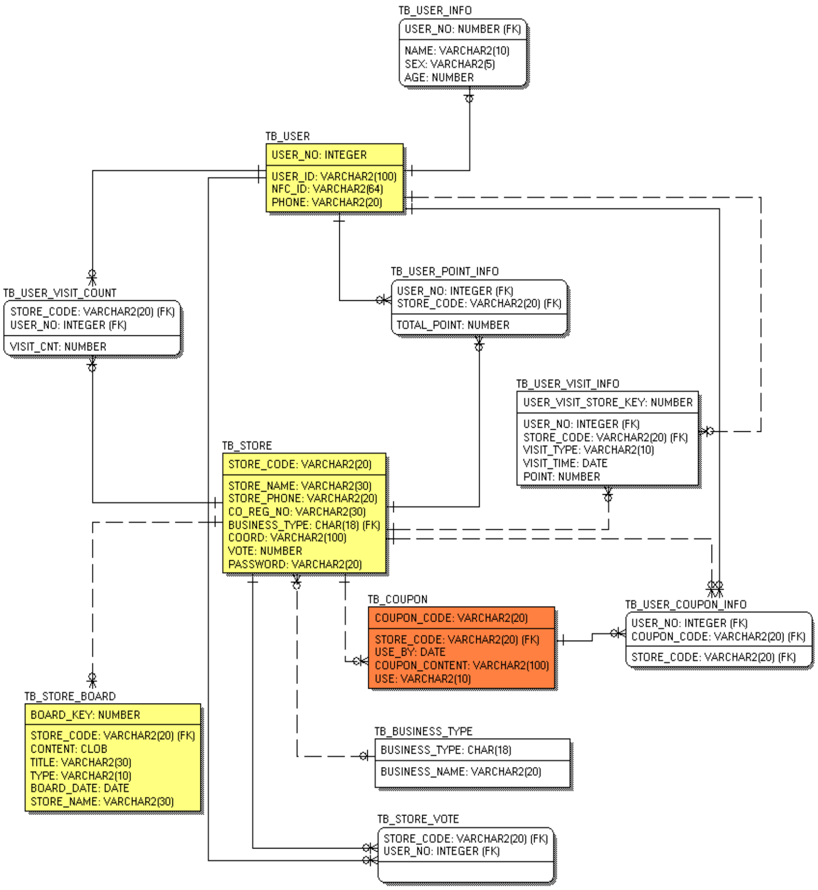

Fig. 3. ER Diagram

Table 2. An example of the Evaluation matrix

	U1	U2	U3	U4	U5
S1	4	3	2		
S2	4	3	2	4	3
S3		2		2	3
S4		3	1		1

For the business visit history, when a customer tags the business's reader using NFC, their usage information will be automatically recorded in the database. It was supposed that if the customer has a favorable view of the business, they would visit it several times; and otherwise the number of visits would be low.

For the nearest neighbor, a network of nearest neighbors is set up for collaborative filtering by measuring the similarity between customers. There are different methods to measure the similarity between customers, including Pearson correlation coefficient, cosine similarity, and Euclidean distance. This study used the Pearson correlation coefficient to measure the similarity between customers because it has a relatively high prediction accuracy.[11] The similarity between customers was measured using the numbers in the review matrix, and then k-NN algorithm was used to set up k nearest neighbors. In Table 2, the estimated customer is U1. Table 3 shows the results when the Pearson coefficient correlation formula is used to calculate the similarity between U1 and other customers.

Table 3. Similarity of customers

U2	U3	U4	U5
0.577	0.9045	0.3015	-0.1924

Therefore, when k, or the number of nearest neighbors, is set to 2, the nearest neighbors of U1 will be U3 and U2. Also, this study set k as one half of the total number of users. For creating a list of recommended items, a method of estimating the scores for items that customers have not visited can be used, based on the data of similar customers. In this study the recommendation list was created using this kind of a method along with the Top-N method.

Table 4. Estimates for items that Customer U1 did not review

	S3	S4
U1	2.718	3.8045

If N is 1 by using the Top-N method, S4, which has the highest estimate between S3 and S4, will be recommended to the user.

4. Implementation of the Mobile Coupon Recommendation System

4.1. Mobile coupon recommendation system scenario

This study supposed that the implemented mobile coupon recommendation system has been in use by customers for a long time. After using the A restaurant, or the business of the system, the customer collects mileage using the NFC functionality on their smartphone with the business's NFC reader. The customer usage data needed by the system can be obtained using this process, and this allows the recommendation system to predict customer usage patterns and tendencies. The customer runs the mobile coupon recommendation app installed on their smartphone and presses the recommendation button from the recommendation page. This recommendation will get sent to the recommendation server, and based on the past usage data of the customer the rules associated with Restruant A will be extracted from the customer's transaction records. The recommendation server will send the extracted results to the customer's smartphone, and the customer can check them in the form of different business types displayed on the screen. The customer selects the next business type to use. Once this is done, the recommendation server predicts the items related to the business type based on similar customers. The recommendation server sends the extracted list of businesses to the customer's smartphone, and the smartphone app sorts the businesses by the nearest straight-line distance using GPS and displays them to the customer. When the customer taps on the business to visit next, they can receive a coupon provided by the business. The customer can manage the coupons as they like: If the customer needs the coupon, they can press the Save button; otherwise they can press the Cancel button.

4.2. Recommendation server

There are two major functions to the recommendation server. The first is extracting related rules from the customer's past usage transactions through association analysis. Fig. 4 shows the association rules extracted from the transactions.

I00001	*I00002*	*0.009708738*
I00001	*I00004*	*0.019417476*
I00001	*I00012*	*0.009708738*
I00001	*I00013*	*0.019417476*
I00001	*I00016*	*0.009708738*
I00002	*I00012*	*0.009708738*
I00002	*I00013*	*0.009708738*
I00002	*I00016*	*0.009708738*
I00004	*I00008*	*0.019417476*
I00009	*I00012*	*0.009708738*
I00012	*I00016*	*0.029126214*

Fig. 4. Association rules for Customer U00023

For the association rule, the one in the form "I#####" shown in Fig. 4 indicates the business type code, while the real number on the right indicates the support level within the transaction. Because arbitrary supposed data were used in this study, the measured support levels for the items were low. Also, with the proposed system the minimum support level in the association analysis was set to 0.001%. The customer used the business type with the business type code I00001 in the past. Therefore, the customer will get recommended the following business types, according to the association rule: I00002, I00004, I00012, I00013, and I00016. Secondly, the customer selects the desired business type, and collaborative filtering is used to recommend the businesses within the business type. Fig. 5 shows setting up the item matrix in the first stage of collaborative filtering, and the finding of similar users in the second stage.

20->44000242122561045670886520457051 1386	
21->102300105406803032042380540706051013	
22->000061020051300806506770423408098083	*20 0.26714908371218610*
23->242001010121400202030400327644400028	*26 0.21675036557652794*
24->033111123404502240302040708050604371	*24 0.17791402751321458*
25->127930101001020801060448060220322010	*21 0.12736273833945494*
26->023000010170100705060500770020242012	*29 0.07724058193556105*
27->101200300705044007407050702402285004	
28->104400979001062307412117210020262041	
29->709040064300104060470530303504032002	

Fig. 5. The item matrix (left) and the similarity with Customer U00023 (right)

The setup of the item matrix, which is the first stage of collaborative filtering, is where the review matrix is set up for all customers within the system. Using the item matrix, and with the k-NN method, similar users were found as shown on the right of Fig. 5. In this study the total number of customers was 10, and k was set to half of that, or 5. Therefore, among the 5 similar customers that were found, the recommendation list was created based on the businesses that Customer U00023 did not visit. Table 5 shows the visit estimates for Customer U00023 using the recommendation list. If a customer selects the business type of I00002, the items that fall under I00002 will be sent to their smartphone.

Table 5. Estimates for items that Customer U1 did not review

Business Type Code	Business Name	Estimation Value
I00013	*Store A*	*1.3137420276347498*
I00013	*Store B*	*1.28444873968987*
I00002	*Chinese Rest. A*	*1.5657812245159438*
I00002	*Chinese Rest. B*	*1.833893624915095*
I00001	*Korean Rest. A*	*-0.256493506493507*
I00001	*Korean Rest. B*	*3.6631330438570986*
I00001	*Korean Rest. C*	*4.3844338069269*
I00004	*Bread A*	*4.055615114299497*
I00001	*Korean Rest. D*	*4.738349847395178*
I00009	*Chicken A*	*3.8863636363636362*
I00012	*Drink A*	*2.037337061834685*
I00001	*Korean Rest. E*	*1.1444653626326835*
I00016	*Coffee A*	*2.1599574810014293*

5. Conclusions

This study implemented the mobile coupon recommendation system, which recommends coupons to the user using association analysis and collaborative filtering based on the consumer usage patterns to solve the problem of information overload. In order to minimize user inconvenience, used data was collected with an implicit method using NFC, and selected the types of businesses with high possibility of user is likely to use by time slot through association analysis. Also, collaborative filtering was used to identify purchase information of similar users for the types of businesses so found, in order to create a recommendation list, centered around businesses that the user has not tried. Through such, both customers and businesses can get recommendation lists that are highly relevant for them.

References

[1] Ericsson Corporate Communications. Global Smart-Phone Subscriptions to Reach 5.6 Billion by 2019. Ericsson Mobility Report; 2013.
[2] Yakup Newspaper. The utilization of Big Data, The success of the Company depends.; 2013.
[3] Min-Jeong Kang. Recommendation System for Make a Purchase to Application in Smart-Phone User. Hongik University Master's Thesis; 2011.
[4] Tae-Ho Hong, Hui-Jeong Lee, Bo-Mil Suh. A Web Personalized Recommender System Using Clustering-based CBR. Korea Intelligent Information Systems Society. Volume 11, No.1, PP.107-121, 2005.
[5] Badrul Sarwar, George Karypis, Joseph Konstan, John Riedl. Item-Based Collaborative Filtering Recommendation Algorithms. In Proceedings of the 10th International World Wide Web Conference, Pages 285-295, 2011.
[6] Korea Communications Agency. Elements according to the step by utilizing big data technologyPropulsion Trends and Implications. Information Communication Technology Issues& Prospects No.10, 2013.
[7] R. Agrawal, T. Imielinski and A. Swami. Mining Association Rules between Sets of Items in Large Database. Proceedings of ACM SIGMOD on Management of Data, PP. 207-216, 1993.
[8] Seong-Hark Kim. A Study on Efficient Searchfor Association Rules and Sequential Patterns. Kon-KuK University Doctoral dissertation, 2004.
[9] Duck-Jin Chai. Mining Association Rules using Transaction Identifiers and Bipartite Grape. Chonnam National University Doctoral dissertation, 2006.
[10] Agrawal, R. and Srikant, R.. Fast algorithms for mining association rules. In Proceeding of the 20th VLDB Conference, pp. 487-499, Santiago, Chile, 1994.
[11] Hee-Choon Lee. Enhancement of Collaborative Filtering in Electronic Commerce Recommender Systems. Kanwon National University Doctoral dissertation, 2009.