Weighted Mining Frequent Itemsets Using FP-Tree Based on RFM for Personalized u-Commerce Recommendation System

Young Sung Cho[1] and Song Chul Moon[2]

[1] Department of Computer Science, Chungbuk National University, Cheongju, Korea
[2] Department of Computer Science, Namseoul University, Cheonan-city, Korea, Korea
youngscho@empal.com, moon@nsu.ac.kr

Abstract. This paper proposes a new weighted mining frequent itemsets using FP-tree based on RFM for personalized u-commerce recommendation system under ubiquitous computing. Existing recommendation system using association rules still does not only reflect exact attributes of item but also has the problem, such as delay of processing speed from a cause of frequent scanning a large data, scalability and accuracy. In this paper, to solve these problems, it is necessary for us to make RFM(Recency, Frequency, Monetary) score of item and to extract the most frequently purchased data from the whole data in order to improve the accuracy of recommendation, to consider frequently changing the weighted patterns by emphasizing the important items with high purchasability according to the threshold for creative the weighted mining frequent itemsets using FP-tree without occurrence of candidate set. To verify improved performance, we make experiments with dataset collected in a cosmetic internet shopping mall.

Keywords: Association Rules, RFM, Weighted Mining Frequent Itemsets using FP-tree.

1 Introduction

Along with the advent of ubiquitous computing environment and the spread of intelligent portable device such as smart phone, PDA and smart pad has been amplified, a variety of services and the amount of information has also increased. It is becoming a part of our common life style that the demands for enjoying the wireless internet are increasing anytime or anyplace without any restriction of time and place[1],[4]. The customers need a recommendation system that can recommend item which they really want on behalf of them. In the u-commerce, it is important to recommend the proper item among large item sets. Therefore, if the recommendation system can recommend the suitable item which they really want, the customers are satisfied with the system. The possession of intelligent recommendation system is becoming the company's business strategy. A personalized recommendation system using data mining technique based on RFM to meet the needs of customers has been actually processed the research[1-5]. We can improve the accuracy of recommendation through an weighted mining frequent itemsets using FP-tree without

James J. (Jong Hyuk) Park et al. (eds.), *Mobile, Ubiquitous, and Intelligent Computing*, Lecture Notes in Electrical Engineering 274,
DOI: 10.1007/978-3-642-40675-1_66,

occurence of candidate set so as to be able to generate the associated items' rules. As a result, we propose a new weighted mining frequent itemsets using FP-tree based on the most frequently purchased data extracted from the whole data for recommendation in u-commerce under ubiquitous computing environment. The next Sect. briefly reviews the literature related to studies. The Sect. 3 is described a new method for personalized recommendation system in detail, such as system architecture with sub modules, the procedure of processing the recommendation, the algorithm for proposing method. The Sect. 4 describes the evaluation of this system in order to prove the criteria of logicality and efficiency through the implementation and the experiment. In Sect. 5, finally it is described the conclusion of paper and further research direction.

2 Related Works

2.1 RFM

RFM segments customers on the basis of how long since they made purchases, how frequently they make purchases, and how much money they spend for segmentation by product usage. The RFM score can be a basis factor how to determine purchasing behavior on the internet shopping mall, is helpful to buy the item which they really want by the personalized recommendation. One well-known commercial approach uses five bins per attributes, which yields 125 cells of segment. The following expression presents RFM score to be able to create an RFM analysis. The RFM score will be shown how to determine the customer as follows, will be used in this paper. The variables (A, B, C) are weights. The categories (R, F, M) have five bins.

$$\text{RFM score} = A \times R \ + \ B \times F \ + \ C \times M \tag{1}$$

The RFM score is correlated to the interest of e-commerce[2]. It is necessary for us to keep the analysis of RFM to be able to reflect the attributes of the item in order to find the items with high purchasability. In this paper, we can use the customers' data and purchased data with 60.98% in the rate of portion for the purchasing counts[4].

2.2 Association Rules

Association rule mining search for interesting relationship among items in a given database. Association rules, first introduced by Agrawal[6], are frequently used by market basket analysis including cross marketing, recommendation system in e-commerce. Association rules which satisfy a minimum confidence threshold are then generated from the frequent itemsets. The traditional association rule mining employs the support measure, which treats every transaction equally. However, in our real world data sets, the weight importance of a pattern may vary frequently due to some unavoidable situations. Usually in an association rules, it is expressed in the form of the rule $X \rightarrow Y$. The rule of $X \rightarrow Y$ means that the transaction including the item of X tends to include the itemsets. And then in an weighted association rules, the w-support of an weighted association rule $X \rightarrow Y$ is defined as

$$\textbf{WSUPP}(\mathbf{X} \rightarrow \mathbf{Y}) = WSUPP(X \cup Y) \quad (2)$$

and the w-confidence is

$$\text{WCONF}(X \rightarrow Y) = \frac{WSUPP(X \cup Y)}{WSUPP(X)} \quad (3)$$

Basically, w-support measures how significantly X and Y appear together; w-confidence measures how strong the rule is. An weighted association rule mining becomes an important research issue in data mining and knowledge discovery by considering different weights for different items. It is necessary to consider these dynamic changes in different application area such as retail market basket data analysis. Much effort has been dedicated to association rule mining with pre-assigned weights[8],[9]. It is crucial to have different weights for different transactions in order to reflect their different importance and adjust the mining results by emphasizing the important transactions.

2.3 Mining Frequent Itemsets Using FP-Tree

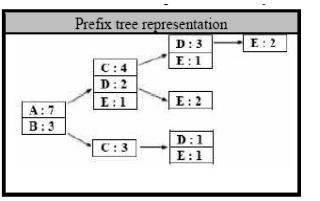
Han et al. [10] proposed a data structure called the FP-tree(Frequent Pattern tree). The FP-tree is a compact representation of all relevant frequency information in a database. Every branch of the FP-tree represents a frequent itemset and the nodes along the branches are stored in decreasing order of frequency of the corresponding items with leaves representing the least frequent items. Compression is achieved by building the tree in such a way that overlapping itemsets share prefixes of the corresponding branches[11]. The FP-tree T has a header table, T:header, associated with it. Single items and their counts are stored in the header table in decreasing order of their frequency. The entry for an item also contains the head of a list that links all the corresponding nodes of the FP-tree. Compared with breadth-first algorithms such as Apriori and its variants, which may need as many database scans as the length of the longest pattern, the FP-growth method only needs two database scans when mining all frequent itemsets. The first scan counts the number of occurrences of each item. The second scan constructs the initial FP-tree which contains all frequency information of the original dataset. Mining the database then becomes mining the FP-tree. To construct the FP-tree, the first scan is to find all frequent items by an initial scan of the database. Then, these items are inserted into the header table in decreasing order of their count. In the next scan, as each transaction is scanned, the set of frequent items in it are inserted into the FP-tree as a branch[11]. If an itemset shares a prefix with an itemset already in the tree, the new itemset will share a prefix of the branch representing that itemset. In addition, a counter is associated with each node in the tree. The counter stores the number of transactions containing the itemset represented by the path from the root to the node in question. This counter is updated during the second scan, when a transaction causes the insertion of a new branch. In this paper, we can use the weighted mining frequent itemsets using FP-tree based on RFM(WMFP) for generating the weighted association rules. In this paper, we can use the weighted mining frequent itemsets using FP-tree based on the customers' data and

Table 1. Example transaction database involving items A,B,C,D, and E

Transaction DB	Lexicographically sorted
{A, D, E}	{A, C, D}
{B, C, D}	{A, C, D, E}
{A, C, E}	{A, C, D, E}
{A, C, D, E}	{A, C, E}
{A, E}	{A, D, E}

Transaction DB	Lexicographically sorted
{A, C, D}	{A, D, E}
{B, C}	{A, E}
{A, C, D, E}	{B, C}
{B, C, E}	{B, C, D}
{A, D, E}	{B, C, E}

Table 2. Prefix tree representation by FP-tree mining



the most frequently purchased data extracted from the whole data to recommend the item they really want exactly.

3 Our Proposal for a Personalized u-Commerce Recommendation System

3.1 System Architecture

We can depict the system configuration concerning the personalized u-commerce recommendation system using the weighted mining frequent itemsets using FP-tree based on RFM under ubiquitous computing environment. This system had four agent modules which have the analytical agent, the recommendation agent, the learning agent, the data mining agent in the internet shopping mall environment. We observed the web standard in the web development, so developed the interface of internet to use full browsing in mobile device. As a matter of course, we can use web browser in wired internet to use our recommendation system. We can use the system under WAP in mobile web environment using feature phone as well as using the internet browser such as safari browser of iPhone and Google chrome browser based on android so as to use our system by using smart phone.

3.2 Weighted Mining Frequent Itemsets Using FP-Tree Based on RFM for Personalized u-Commerce Recommendation System

In this part, we can depict the weighted mining frequent itemsets using FP-tree based on RFM for personalized u-commerce recommendation system. Our algorithm can consider situation where the weight / importance of a pattern may vary dynamically in e-commerce on the real world. It is necessary for us to consider the quantity of purchased data extracted by the scope of RFM score which is between the score is more than 19 points and the score is less than 40 points, had a lot of purchasing counts in order to prevent delay of processing speed from a cause of frequent scanning a large data. We can calculate the rate of weight based on the quantity item by each rank of the RFM score. We can have different weights for different transactions and to generate the weighted association rules through the weighted mining frequent itemsets using FP-tree based on RFM. Thus we can forecast frequently changing trends by emphasizing

the important items. At first, we can aggregate the quantity of purchased data(sale_dat3) extracted from the whole data (sale). After that, we can make the rate of weight using aggregated counts of a section, that is, it is become the value of weight based on the quantity of extracted data. To propose our paper, we have two step of procedural preprocessing. The 1st step is that it is necessary for us to make the RFM scores to reflect the attributes of the item. The 2nd step is that it is necessary for us to process an weighted mining frequent itemsets using FP-tree based on the most frequently purchased data extracted from the whole data in order to forecast frequently changing trends by emphasizing the important items with high purchasability according to the threshold for creative the weighted association rules.for generating association rules with w-support, w-confidence and w-lift through the weighted mining frequent itemsets using FP-tree based on RFM(WMFP). The procedural algorithm of WMFP for personalized u-commerce recommendation system is depicted as the following Table 3.

Table 3. The procedural algorithm of WMFP for personalized u-commerce recommendation system

Step 1 : The RFM score of customer is computed so as to reflect the attributes of the customer, consists of three attributes(R, F, M), each attribute has five bins divided by each 20%, exact quintile.
Step 2 : The system can aggregate the quantity of purchased data by each interval customer's RFM scores, which is aggregated counts of distribution from the whole data, make the rate of weight.
Step 3 : The system can calculate the rate of weight based on quantity item with each rank of RFM score for customer.
Step 4 : The system can scan extracted database(sale_dat3) and make the weighted mining frequent itemsets using FP-tree based on sale_dat3.
Step 5 : Weighted Association rules are created by WMFP
Step 6 : Wsupport = $\sum^{n}$ *Weight/N X Support count /* N is numbers of item in the rules */*
Step 7 : The system *can create creative the weighted association rules with w- support, w-confidence and w-lift through the weighted mining frequent itemsets using FP-tree using FP-tree without occurrence of candidate set.*
Step 8 : The system can reflect frequently changing the weighted patterns by emphasizing the important items.
Step9: The system can recommend the items with high purchasability according to the threshold for creative the weighted association rules with w-support, w-confidence and w-lift through WMFP.

3.3 The Procedural Algorithm for Personalized u-Commerce Recommendation System

The system can search the information in the cluster selected by using the code of classification and customer's RFM score in users' information. It can scan the preference as the average of brand item in the cluster, suggest the brand item in item category selected by the highest probability for preference as the average of brand item. This system can create the list of recommendation with TOP-N of the highest preference of item to recommend the item with purchasability efficiently. This system can recommend the items with efficiency, are used to generate the recommendable item according to the basic threshold for the weighted mining frequent itemsets using FP-tree, with w-support, w-confidence and w-lift. It can recommend the associated item to TOP-N of recommending list if users want to have the cross-selling or up-selling. This system takes the cross comparison with purchased data in order to avoid the duplicated recommendation which it has ever taken.

4 The Environment of Implementation and Experiment and Evaluation

4.1 Experimental Environment

This system proposes a new weighted mining frequent itemsets using FP-tree based on RFM under ubiquitous computing environment. In order to do that, we make the implementation for prototyping of the internet shopping mall which handles the cosmetics professionally and do the experiment. It is the environment of implementation and experiment below.

- OS: Windows XP SP2,
- Web Server: Apache 2.2.14 / WAP 2.0
- Server-Side Script : JSP/PHP 5.2.12
- XML/WML2.0/ HTML5.0/CSS3/JAVASCRIPT
- Database: MySQL 5.1.39
- J2SDK(1.7.0_11)
- MySQL JDBC
- J2SDK(1.7.0_11)
- jakarta-tomcat (5.0.28)

We have carried out the implementation and the experiment for proposing system through system design, we have finished the system implementation about prototyping recommendation system. It could be improved and evaluated to new system through the result of experiment with the metrics such as precision, recall, F-measure as comparing proposing ystem(W_FP) with other previous system(O_ARM) using original method of mining and existing system.

4.2 Experimental Data for Evaluation

We used 319 users who have had the experience to buy items in e-shopping mall, 580 cosmetic items used in current industry, 1600 results of purchase data recommended in order to evaluate the proposing system[4]. In order to do that, we make the implementation for prototyping of the internet shopping mall which handles the cosmetics professionally and do the experiment. We have finished the system implementation about prototyping recommendation system. We'd try to carry out the experiments in the same condition with dataset collected in a cosmetic internet shopping mall. It could be evaluated in MAE and Precision, Recall, F-measure for the recommendation system in clusters. It could be proved by the experiment through the experiment with learning data set for 12 months, testing data set for 3 months in a cosmetic cyber shopping mall[4]. The 1st system of the weighted mining association rules based quantity item with RFM score, is proposing method(W_FP) called by "proposal", the 2nd system is the original method(O_ARM) using the ordinary association rules mining, the third system is existing system. The proposing method's overall performance evaluation for recommendation is precision, recall and F-measure as comparing proposing method using (W_FP) and the original method using (O_ARM). The performance was performed to prove the validity of recommendation and the system's overall performance evaluation. The metrics of evaluation for recommendation system in our system was used in the field of information retrieval commonly[12].

Table 4. The result for table of precision, recall, F-measure for recommendation ratio by each cluster

Cluster	Proposal(W_FP)			O_ARM			Existing		
	Precision1	Recall1	F-measure1	Precision2	Recall2	F-measure2	Precision3	Recall3	F-measure3
C1	42.23	80.66	55.44	46.20	55.70	45.90	48.79	31.32	35.64
C2	45.25	84.47	58.93	43.20	52.53	45.76	49.36	29.54	35.06
C3	46.42	76.94	57.90	40.99	23.93	30.05	44.26	21.81	27.65
C4	42.81	64.52	51.47	56.06	38.37	43.77	52.49	34.98	39.75
C5	54.99	78.45	64.66	44.83	47.40	42.74	50.41	43.21	43.10
C6	48.56	81.40	60.83	33.56	37.23	34.68	50.93	36.60	39.64
C7	41.58	74.08	53.26	53.94	27.27	34.90	47.41	26.81	32.26
C8	53.92	78.87	64.05	45.07	37.23	38.29	43.60	36.60	37.82
C9	27.72	74.08	40.34	64.08	28.45	37.19	46.68	25.19	30.28
C10	59.99	93.78	73.17	73.85	62.50	64.42	67.23	55.34	57.10

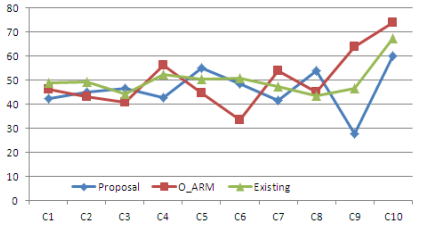


Fig. 1. The result of recommending ratio by precision

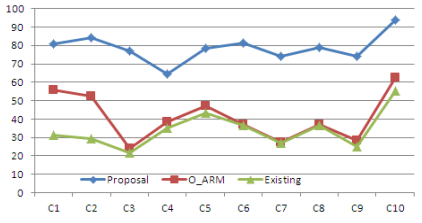


Fig. 2. The result of recommending ratio by recall

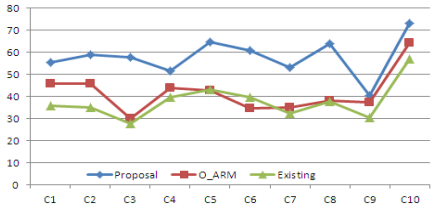


Fig. 3. The result of recommending ratio by F-measure

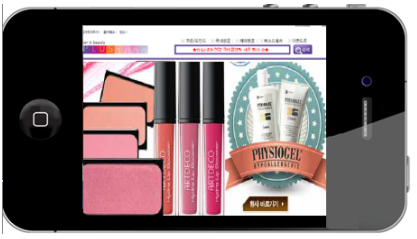


Fig. 4. The result of recommending items of cosmetics

Above Table 4 presents the result of evaluation metrics (precision, recall and F-measure) for recommendation system. The weighted mining frequent itemsets using FP-tree is improved better performance of proposing method using (W_FP) than the original method using (O_ARM). The proposed higher 37.66% in recall even if it is lower 3.83% in precision than the original method using (O_ARM), higher 16.24% in F-measure than the original method. After that, it shows that our algorithm is very efficient and scalable for the recommendation system. Above figure 4 is shown in the result of screen on a smart phone. The performance of proposing mining method was improved more counts of support and rule than the original method, it was especially worthy of notice, in the rule counts, had an effect about 4 times what the original mining method did before. As a result, it was efficient for us to recommend the items of association because it is strong cohesion of the attribute of item based the weighted association rules using the weight based quantity item with RFM score. So, we could have the recommendation system to be able to recommend the items with high purchasability. Above figure 4 is shown in the result of screen on a smart phone. The performance of proposing method is improved although it is less in average of confidence (average confi_rate), however it is efficient for us to recommend the items of association because it is strong cohesion of the attribute of item because of using a new weighted mining frequent itemsets using FP-tree based on RFM.

5 Conclusion

Recently u-commerce as a application field under ubiquitous computing environment, is in the limelight[5]. Existing recommendation system using association rules still does not only reflect exact attributes of item but also has the problem, such as delay of processing speed from a cause of frequent scanning a large data, scalability and accuracy. And also, existing algorithms for mining association rules are based on fixed weight, do not reflect the weight / importance of a pattern, and do not consider these dynamic changes in different application area such as retail market basket data analysis. It was necessary for us to make RFM score of item and to extract the most frequently purchased data from the whole data in order to improve the accuracy of recommendation, to reflect frequently changing the weighted patterns by emphasizing the important items with high purchasability according to the threshold for creative the weighted mining frequent itemsets using FP-tree without occurrence of candidate set in order to solve these problems, to use the dynamic weights in proposing method of mining. We could improve the performance of the weighted mining rapidly. As a result, we proposed a new weighted mining technique using FP-tree for personalized u-commerce recommendation system in real datasets environment in order to improve the accuracy of recommendation with high purchasability. As a matter of course, we have described that the performance of the proposing method with the weighted mining using FP-tree based RFM was improved better than the original method and existing system in mining test. It is meaningful to present a new weighted mining technique using FP-tree based on quantity item with RFM score for personalized u-commerce recommendation system under ubiquitous computing environment. The following research will be looking for a personalized recommendation in semantic web environment by fuzzy clustering approach to increase the efficiency and scalability under ubiquitous computing environment**.**

Acknowledgements. This paper[2)] was supported by funding of Namseoul University.

References

1. Cho, Y.S., Ryu, K.H.: Personalized Recommendation System using FP-tree Mining based on RFM. In: KSCI, vol. 17(2) (February 2012)
2. Jin, B.W., Cho, Y.S., Ryu, K.H.: Personalized e-Commerce Recommendation System using RFM method and Association Rules. In: KSCI, vol. 15(12), pp. 227–235 (December 2010)
3. Cho, Y.S., Moon, S.C., Noh, S.C., Ryu, K.H.: Implementation of Personalized recommendation System using k-means Clustering of Item Category based on RFM. In: 2012 IEEE International Conference on Management of Innovation & Technology Publication (June 2012)
4. Cho, Y.S., Moon, S.C., Jeong, S.P., Oh, I.B., Ryu, K.H.: Clustering Method using Item Preference based on RFM for Recommendation System in u-Commerce. In: Han, Y.-H., Park, D.-S., Jia, W., Yeo, S.-S. (eds.) Ubiquitous Information Technologies and Applications. LNEE, vol. 214, pp. 353–362. Springer, Heidelberg (2012)

5. Cho, Y.S., Moon, S.C., Ryu, K.H.: Mining Association Rules using RFM Scoring Method for Personalized u-Commerce Recommendation System in Emerging Data. In: Kim, T.-H., Ramos, C., Abawajy, J., Kang, B.-H., Ślęzak, D., Adeli, H. (eds.) MAS/ASNT 2012. CCIS, vol. 341, pp. 190–198. Springer, Heidelberg (2012)
6. Agrawal, R., Imielinski, T., Swami, A.: Mining Association Rules between Sets of Items in Large Datasets. In: Proc. ACM SIGMOD 1993, pp. 207–216 (1993)
7. Agrawal, A., Faloutsos, C., Swami, A.: Mining association rules between sets of items in large databases. In: Proceedings of the ACM SIGMOD International Conference on Management of Data, Washington, D.C., pp. 207–216 (May 1993)
8. Ramkumar, G.D., Ranka, S., Tsur, S.: Weighted Association Rules: Model and Algorithm. In: Proc. ACM SIGKDD (1998)
9. Tao, F., Murtagh, F., Farid, M.: Weighted Association Rule Mining Using Weighted Support and Significance Framework. In: Proc. ACM SIGKDD 2003, pp. 661–666 (2003)
10. Han, J., Pei, J., Yin, Y., Mao, R.: Mining Frequent Patterns without Candidate Generation: A Frequent-Pattern Tree Approach. Data Mining and Knowledge Discovery 8(1), 53–87 (2004)
11. Grahne, G., Zhu, J.: Efficiently Using Prefix-trees in Mining Frequent Itemsets. In: Proc. IEEE ICDM 2003 Workshop FIMI 2003. CEUR Workshop Proc. series, vol. 90 (2003), http://ceur-ws.org/vol-90
12. Herlocker, J.L., Kosran, J.A., Borchers, A., Riedl, J.: An Algorithm Framework for Performing Collaborative Filtering. In: Proceedings of the 1999 Conference on Research and Development in Information Research and Development in Information Retrival (1999)