

Finding Frequent Itemsets of Healthy Shop Products Using APRIORI Algorithm

Ngwe Zin Khaing, Khin Sandar, May Phyo Oo
Computer University, Patheingyi
ngwezinkhine@gmail.com, drkhinsandar@gmail.com, mayphyooo@gmail.com

ABSTRACT

Data mining in an area in the intersection of machine learning statistics, and database is to use searching for relationships and global pattern that exists, but is hidden in large database. The discovery of interesting association relationships among huge amount of business transaction records can help in many business decision making processes, such as catalog design, cross marketing and loss leader analysis. This paper intends to an effective data mining process that contains the concept of market basket analysis using association rule mining. Apriori is an influential algorithm for mining frequent itemsets under Boolean association rules. The name of the algorithm is based on the fact that the algorithm uses prior knowledge of frequent itemsets properties. This system analyzes the customer buying habits of healthy shop products and finding associations between the different items that customer places in their shopping baskets. And their rules are generating from frequent item sets based on the healthy shop transaction data. The result is Market Basket Analysis in web interface that made easier to navigate and visualize the data.

Keywords: Apriori Algorithm, Frequent Itemsets, Association Rule, Primary/Foreign Keys, finding frequent itemsets mining

1. INTRODUCTION

Data mining is the process of analysis of raw data in the database and synthesizing it into information that is useful for effective decision making. Data mining automates the process of finding relationships and patterns in raw data and delivers results that can be either utilized in an automated decision support system or assessed by a human analyst [2].

Data mining can be done most easily on data that has been moved or processed into a data warehouse. Mining does not mean analyzing details of data alone. General data mining tools are also beginning to appear, without a specific business market but relying on defined relationships in the

data model. The neural network tools have had a tremendous impact on businesses when searching for information to solve problems that human analyst could not find. What has happened with data mining is that technology has finally leaped into an information processing arena from the order daily Data can be stored in many different types of databases. On Line Analytical Processing (OLAP) is an analysis technique with functionalities such as summarization, consolidation and aggregation, as well as the ability to view information from different angles [6]. In recent years, companies often realize that the disparity of data in pre-existing or legacy systems makes accurate data integration virtually impossible.

The Various forms of electronic marketplaces – auctions, product exchanges, online shopping market, e-catalogs, etc. - are considered for critical elements of electronic commerce. Concept is Virtual Exhibitions. Virtual Shopping Malls capture this dimension of virtual. These do not physically exist. The healthy shop productions have made the customers feel pampered and beautiful. They production have treated their complexion with understanding and the science of beautiful pure nature is an enormous benefit [4]. So the healthy shop realized the truth of nature. This can be brought to use in the manufacture of products for the highest benefit. In addition, with state-of-the-art scientific technology and the selection of product from natural sources, we can promise sincerely and whole-heartedly that we manufacture these products to protect the customers' skin, care for beauty and meet their demands always. In this paper, related work is described in section (2), proposed system in section (3), implementation of the system in section (4) and conclusion in section (5).

2. RELATED WORK

Many businesses are using data warehouses which have been built from operational data housing for all the details of the business. The larger problem is not the accessing of that data, but whether the data is clean. This adds a layer of complexity to the data mining warehouse, the validation of the data being searched. Many redundancies will exist in the data. These will not be easily identified because non-related application areas use the same data but call it differently.

Analytical processing will provide an effective platform for data mining [4]. Most future database management system will be based on objects or at least will incorporate management object oriented functionally [5]. The data is loading into the target database by using SQL statement [4]. A formal definition of the data warehouse id offered by W.H.Inmon:“A data warehouse is a subject-oriented, integrated, time-variant, and nonvolatile”.

“Collection of data in supports to make management decisions” Association rule mining raised by Rakesh Agrawal is an important research problem in data mining field [6]. Association rule mining aims at detecting the relationship of tuples in transactional database and serving decision making [1]. The association rule is the pattern knowledge existing in the given dataset or database. Let’s data itemset $X = \{X_1, X_2, \dots, X_m\}$, $Y = \{Y_1, Y_2, \dots, Y_n\}$, association rule can be represented as the form of “ $X \rightarrow Y$ ”, like “ $X_1 \cap X_2 \cap \dots \cap X_m \rightarrow Y_1 \cap Y_2 \cap \dots \cap Y_n$ ”; $X_i (i \in \{1, \dots, m\})$ and $Y_j (j \in \{1, \dots, n\})$ are all in attribute-value format, and $X \cap Y = \Phi$. Association rule $X \rightarrow Y$ expresses “tuples satisfying conditions in X also satisfy conditions in Y”(Zhu Ming, 2002). If rule $X \rightarrow Y$ is true, then it has the support and confidence values. The name of the algorithm is based on the fact the algorithm uses prior knowledge of frequent item set properties. A two-steps process is followed, consisting of join and prune actions. The results is overall sales frequent item set.

3. PROPOSED SYSTEM

Overview of proposed system Architecture and System Flow Diagram are described in figure 1 and 2. This system intends to an effective data mining process that contains the concept of market basket analysis using association rule mining. The result is Market Basket Analysis in web interface that made easier to navigate and visualize the data. This system has two main parts which are customer shopping and administration of the system. The customer shopping system displays the information of sale item. It includes the flow of required steps of customer shopping. In Administration, an analyst or manager can analyze the over all sales reports of the market. Especially, the system display overall sales frequent item sets which are greater than or equal minimum support value analyzed by the manager or analyst.

Market basket analysis is implemented with the selling items of online shopping. The healthy shop nature is just buying items by a customer from Healthy Shop Product. The customer can select any item which he/she wants to buy and add to his/her shopping cart. And then the customer can remove or buy any more items as much as he/she wants to buy before making checkout. Market basket analysis is

used to determine which products sell together on each of transaction.

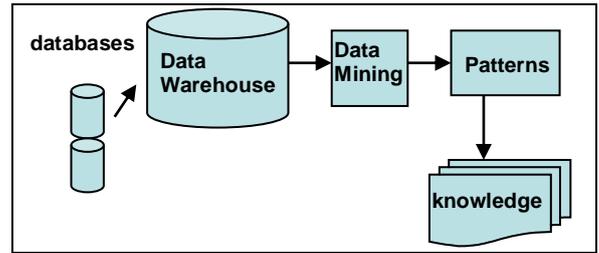


Figure 1. Overview of the System Architecture

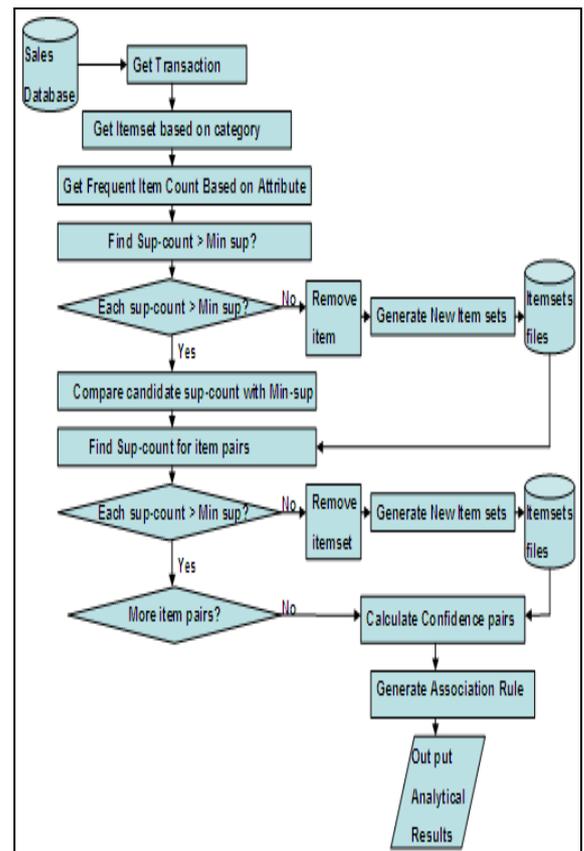


Figure 2. The System Flow Diagram

In this system, Sales data is used to transaction data by using apriori algorithm and finding frequent itemsets. Attributes are used to the transaction data.

Table 1. Transaction Data

TID	Items
1	HP 116, HP 104, HP 221
2	HP 112, HP 104, HP 508
3	HP 116, HP 112, HP 104, HP 508
4	HP112, HP 508

Code No.	Description	Code No.	Description
HP100	Anti-Wrinkle Cream	HP210	Hand Wash
HP101	Day Cream	HP211	Hand Cream
HP102	Lift -up Face & Neck Cream	HP214	Pink Nipple
HP103	Eye Contour Gel	HP221	Sun Block Face
HP104	White Cream	HP222	Sun Block Body
HP105	Anti Acne Cream	HP226	Nail Polish
HP106	Brightening Serum	HP227	Propolis Soap
HP107	Double Action Collagen	HP229	Body Milky Ice Cream
HP108	Oil Control Gel	HP504	Eye Shadow
HP109	Face Mud Cream	HP506	Eye Liner
HP110	Face Scrub	HP508	Exclusive Compact Powder
HP112	Cleanser Cream	HP510	Loose Powder
HP116	Facial Foam	HP512	Liquid Powder

Figure 3. Attributes Used in the System

In this system represents the operation data store. It gets the data from daily sales or transactional data. In this system, we existing Apriori algorithm.

Algorithm: Apriori. Find frequent itemsets using an iterative level wise approach based on candidate generation.

Input: Database, D, of transaction; minimum support threshold, min_sup.

Output: L, frequent itemsets in D.

Method:

1. $L_1 = \text{find_frequent_1-itemsets}(D)$;
2. for $(k=2; L_{k-1} \neq \Phi; k++)$ {
3. $C_k = \text{apriori_gen}(L_{k-1}, \text{min_sup})$;
4. for each transaction $t \in D$
5. $C_t = \text{subset}(C_k, t)$;
6. for each candidate $c \in C_t$
7. $c.\text{count}++$
8. }
9. $L_k = \{c \in C_k \mid c.\text{count} \geq \text{min_sup}\}$
10. }
11. return $L = \bigcup_k L_k$;

Procedure apriori_gen (L_{k-1} : frequent (k-1)-itemsets; min_sup: minimum support thresholds)

1. for each itemset $L_1 \in L_{k-1}$
2. for each itemset $L_2 \in L_{k-1}$
3. if $(l_1[1] = l_2[1] \wedge l_1[2] = l_2[2] \wedge \dots \wedge l_1[k-2] = l_2[k-2] \wedge \dots \wedge l_1[k-1] < l_2[k-1])$ then {
4. $c = l_1 \cup l_2$
5. if has infrequent subset(c, L_{k-1}) then delete c;
6. else add c to C_k ;

8. }
9. return C_k ;

Procedure has_infrequent_subset (c: candidate k-itemset; L_{k-1} : frequent (k-1)-Itemsets);

1. for each (k-1)-subset s of c
2. if $s \in L_{k-1}$ then
3. return TRUE;
4. return FALSE;

4. IMPLEMENTATION OF THE SYSTEM

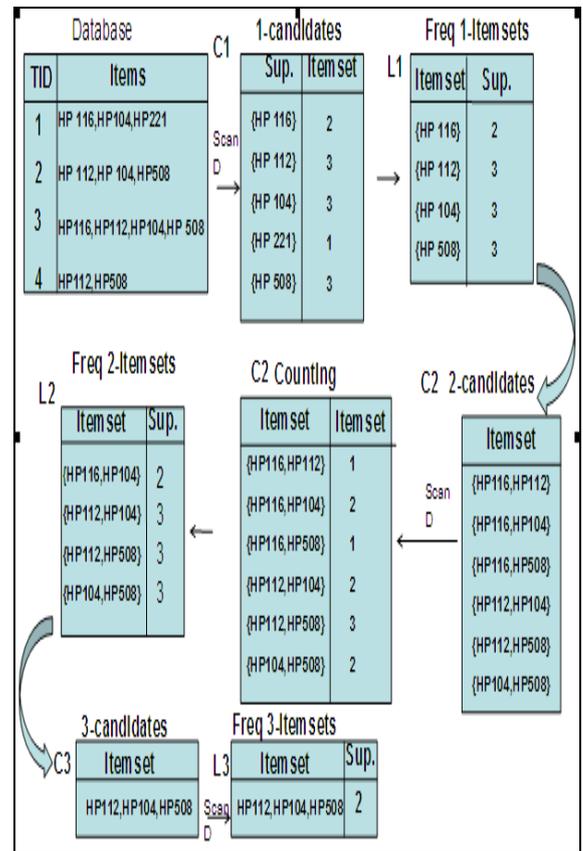


Figure 4. Generation of Candidate Itemsets and Frequent Itemsets, Where the Minimum Support Count is 2

The system is implemented Market Basket Analysis System. It may be performed on the detail data transaction of customers on the market. The result may be used to plan of marketing, advertising strategies as well as catalogue design and store layout of selling items. In another strategy, items that are frequently purchased together can be placed in close proximity in order to further encourage the sale of such items together.

According to figure 4, this system shows Apriori algorithm and its related procedures. (Step 1) of apriori finds the frequent 1- itemsets, L_1 . In steps (2-10), L_{k-1} is used to generate candidates C_k

in order to find L_k . The apriori_gen procedure generates the candidates and then uses the Apriori property to eliminate those having a subset that is not frequent (Step 3). The procedure is described below. Once all the candidates have been generated, the database is scanned (Step 4). For each transaction, a subset function is used to find all subsets of the transaction that are candidates (Step 5), and the count for each of these candidates is accumulated (step 6 and 7). Finally, all those candidates satisfying minimum support form the set of frequent itemsets, L . A procedure can then be called to generate association rules from the frequent itemsets.

The apriori_gen producer performs two kinds of actions, join and prune. In the join component, L_{k-1} is joined with L_{k-1} to generate potential candidates (Step 1-4). The prune component (Step 5-7) employs the Apriori property to remove candidates that have a subset that is not frequent. The test for infrequent subsets is shown in procedure has_infrequent_subset. This system includes a banner pane of animated advertising on the top of the page, the main menu button are on the left side of the page and the main home page. There are home eight menu buttons:

1.Home 2.Register 3.Login 4.Categories
5.View Card 6.Checkout 7.Default 8.Frequent Item.
There are eight menus to analyze the required information of the entire market. The following is the overall sales of the entire market and overall sales of the administrator. The following figure result is overall sales of the market.

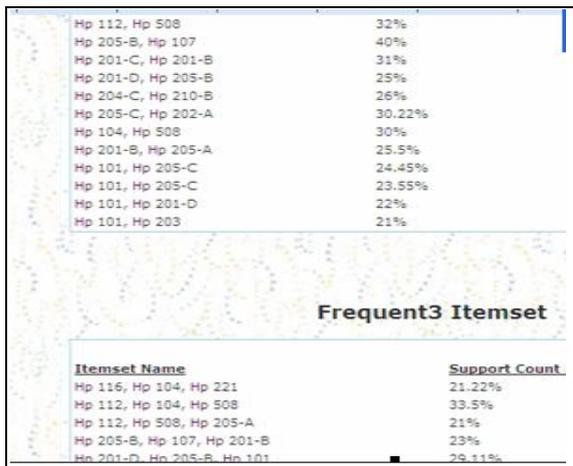


Figure 5. Overall Sales Frequent Itemset

5. CONCLUSION

This system has been implemented by using Active Server Page (ASP) programming language. This system intends to an effective data mining process that contains the concept of market basket analysis using association rule mining. Hence. Two

or more items can be bought together, by using this pattern, often from the market basket data, where the base information consist of register transaction of stores. Market Basket Analysis or Manger can be supported to know habits of the customers and which item groups set are more purchased together in a single transaction due to good results of this system.

REFERENCES

- [1] Jeremy Mennis, Jun Wei Liu, 2005. Mining Association Rules in Spatio-Temporal Data. Transactions in GIS, 9(1): pp. 5-17
- [2] Zhu Ming, 2002. Data Mining. University of Science and Technology of China Press, Hefei, pp. 11-12
- [3] Han Jiawei and Kamber Michelin Data Mining: Concepts and Techniques ISBN 1-55860-489-8
- [4] Using Automate for Data Warehousing <http://.bbk.ac.UK>
- [5] Modern Database Management Sixth Edition
- [6] Berson Alex and J. Smith Stephen Data Warehousing, Data Mining & OLAP ISBN0-07-120257-91997byTheMcGraw-Hill Companies, Inc
- [7] Bischoff Joyce and Alexander Ted Data Warehousing: Practical Advice from the Experts ISBN 0-13-577370-9 1997 by Prentice Hall, Inc
- [8] J. Han, M. Kamber(2001), Data Mining, Morgan Kaufmann Publishers, San Francisco, CA
- [9] Data Warehousing Concept and Mechanism www.dw-instrute.com
- [10] Ferguson, Mike Evaluating and Selecting Data Mining Tools 2001 by Academic Press
- [11] Migration to Open Source Database <http://www.osdb.migration.org>
- [12] Extension In Data Warehousing <http://.www.stand.ford.edu.org>
- [13] Date C.J. An Introduction to Database System ISBN 0-201-68419-5 2000 by Addison Wesley Longman, Inc (Seventh Edition)
- [14] Healthy Shop Products Web Site <http://www.healthysshop-nature.com>