# Association Rule Based Data Mining System for Sale System

Ni Ni Aung
*Computer University, (Lashio)*
*nangnyinyilay@gmail.com*

## Abstract

*Computer software are widely used in economic functions. It has sale system, accessory system and rate system. There are so many kinds of analysis tools to find out number of sale, list of credit and loss and profit. Nowadays, many industries are becoming interested in mining association rule from their database with a large amount of data being collected and stored. Data Mining is the process of discovering interesting patterns from large amount of data in database. In this paper, the best sale itemsets are searched based on data mining system. The frequent itemsets are searched by using Apriori Algorithm and then produce the Association Rule. Frequent itemsets are the best sale itemsets and to find them, user must input the minimum support count. Then, the manager or retailers can use these results for planning marketing or advertising strategies, catalog design, as well as different store layouts*.

## 1. Introduction

With the growing importance of businesses and analyze of these, there is also a growing interest in tool that can analyze the sale information. Data mining tools perform data analysis and many uncover important data patterns, contributing greatly to business strategies, knowledge bases, and scientific and medical research. Data Mining (sometimes called data or knowledge discovery) is the process of analyzing data from different perspectives and summarizing it into useful information. Information that can be used to increase revenue, cuts costs, or both. Data mining software is one of a number of analytical tools for analyzing data. Such analysis can support to make decision about marketing, management, inventory, promotion, etc. Mining finds interesting association or correlation relationships among a large of data items.

The Association Rule function is often associated with “Market Basket Analysis”. Market Basket Analysis is a technique that assists in understanding what items are likely to purchase together according to the association rules, primarily with the aim of identifying cross-selling opportunities. Knowing and analyzing what products people purchases as a group can be very helpful to a retailer in particular or to any other seller in general. A shopping center can use this technique to organize and place products frequently sols together into the same area.

## 2. Related Work

The works related to the proposed system are presented here. The system is based data mining field. Therefore, the details of these techniques are presented. Data mining also known as Knowledge Discovery in Databases (KDD), is to find trends, patterns, correlations, anomalies, in these databases which can help us to make accurate future decisions. A general view of data mining can be found in [4, 6].The aim of data mining to find novel, interesting and useful patterns from data using algorithms (methods of finding such information).

J.Han and M.Kamber [3] describe the Association Rule Mining, Market Basket Analysis, Boolean association rule, Apriori Algorithm and how to mine the data from the database. Sergey Brin, Rajeev Montwani, Jeffery D.Ullman and Shalom Tsur [10] also describe itemsets counting and implication rules for market basket data. Moreover C.C. Aggarwal and P.S.Yu. described new framework for itemsets generation. Edward R.Omiecinski [2] presented the alternative interest measures for mining association in database. Rakesh Agrawal and Ramakrishnan Srikant [8] consider fast algorithms for mining association rules in large databases.

## 3. Data Mining

Data Mining is the process of discovering interesting knowledge from large amounts of data stored either in databases, data warehouses, or other information repositories. Data mining has attracted a great deal of attention in the information industry in recent years is due to the wide availability of huge amounts of data and the imminent need for turning such data into useful information and knowledge. The information and knowledge gained can be used for application ranging from business management, production control, and market analysis, to engineer design and science exploration. Data mining can be viewed as a result of the natural evolution of information technology.

### 3.1. Market basket analysis

The Association Rules function is often associated with “market basket analysis”, which is used to discover relationships or correlations among a set of items. This process analyzes customer buying habits by finding associations between the different items that customers place in their “shopping baskets”. The discovery of such associations can help retailers develop marketing strategies by gaining insight into which items are frequently purchased together by customers. Such information can lead to increased sales by helping retailers do selective marketing and plan their shelf space.

### 3.2. Association Rule

Association Analysis is used to discover itemsets that describe strongly associated features in the items. Rules that satisfy both a minimum support threshold (min_sup) and a minimum confidence threshold (min_ conf) are called strong. The discovery of interesting association relationships among huge amounts of business transaction records can help in many business decision making processes, such as catalog design, cross marketing, and loss- leader analysis [3]. Association Rule Mining is a two step process:

1. Find all frequent itemsets : By definition, each of these itemsets occur at least as frequently as a pre-determined minimum support count.
2. Generate strong association rules from the frequent itemsets: By definition, these rules must satisfy minimum support and minimum confidence.

## 4. Mining steps of frequent itemsets

The next is mining the frequent itemsets from the transaction database. The Apriori Algorithm are implemented which is one of the first and fastest methods. Most of the methods for Mining Association Rules depend on this algorithm.

The Apriori Algorithm uses a bottom-up breadth-first approach to find the large itemsets. The algorithm Apriori works as follows:

It first generates the one itemsets that have support greater than a pre specified minimum support count. This task is done for each item counting the number of occurrences and selecting those whose support is greater than minimum support count. Then the procedure generates two itemsets using these large one itemsets with the procedure Apriori_Gen. There is a pruning step which prunes the generated two itemsets. The algorithm goes on generating the next itemsets and pruning until no large itemsets is left [3].

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

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

Output : L, frequent itemsets in D.

Method :

(1) L1 = find _ frequent _ 1-itemsets (D);
(2) for (k = 2;$L_{k-1} \neq \emptyset$;k++) {
(3) $C_k$ = apriori_gen ($L_{k-1}$, min_sup);
(4) For each transaction $t \in D$ {// scan D for counts
(5) $C_t$ = subset ($C_k$, t); // get the subset of t that are candidates
(6) for each candidate $c \in Ct$
(7) c.count ++;
(8) $L_k = \{c \in C_k \mid c.count \geq min\_sup\}$
(9) return $L = U_k L_k$;

Procedure apriori_gen

($L_{k-1}$: frequent (k-1)- itemsets; min_sup: minimum support threshold )

(1) for each itemset $l_1 \in L_{k-1}$
(2) for each itemset $l_2 \in L_{k-1}$

(3) if ( $l_1[1] = l_2[1]$ ) $\wedge$ ($l_1[2] = l_1[2]$ ) $\wedge$….
$\wedge$
($l_1[k-2]=l_2[k-2]$ )$\wedge$($l_1[k-1]< l_2[k-1]$) then {

(4) c = $l_1 \infty l_2$:// join step: generate candidate

(5) if has_ infrequent _subset(c, $L_{k-1}$) then

(6) delete c;//prune step: remove unfruitful candidate

(7) else add c to $C_k$;

(8) }

(9) return $C_k$;

procedure has_ infrequent _subset (c; candidate k-itemset; Lk-1: frequent ( k-1) – itemsets);

(1) for each (k-1) – subset of c

(2) if s $\notin$ $L_{k-1}$ then

(3) return TRUE ;

(4) return FALSE;

A transaction database (TDB) consists of a set of transactions in the form of D = (TID, list of item) where TID is a transaction_id. We use an example with the transaction database shown in Table4.

| **Table 4** | |
|---|---|
| TID | Itemlist |
| T100 | Pen, Pencil, Ruler |
| T200 | Pencil, Correctionpen |
| T300 | Pencil, Eraser |
| T400 | Pen, Pencil, Correctionpen |
| T500 | Pen, Eraser |
| T600 | Pencil, Eraser |
| T700 | Pen, Eraser |
| T800 | Pen, Pencil ,Eraser, Correctionpen |
| T900 | Pen, Pencil, Eraser |

At first, all the items are scanned in order to count the number of occurrences of each term. The set of candidate 1- itemsets is shown in Table 4.1.

| **Table 4.1** | |
|---|---|
| Itemsets | sup_count |
| Pen | 6 |
| Pencil | 7 |
| Eraser | 6 |
| Correctionpen | 2 |
| Ruler | 2 |

In this example, the minimum support count is set with 2.Then candidate support count are compared with minimum support count. The set of frequent 1- itemsets, L1, is shown in Table 4.2. It consists of the candidate1- itemset satisfying minimum support count.

| **Table 4.2** | |
|---|---|
| Itemsets | sup_count |
| Pen | 6 |
| Pencil | 7 |
| Eraser | 6 |
| Correctionpen | 2 |
| Ruler | 2 |

To discover the set of frequent 2- itemsets, L2, L1 with L1 are joined to generate a candidate 2-itemsets, C2. It consists of 2-itemsets is described in Table 4.3. Then transaction database is scanned for count of each candidate. The result of C2 is shown in Table 4.4.

| **Table 4.3** |
|---|
| Pen, Pencil |
| Pen, Eraser |
| Pen, Correctionpen |
| Pen, Ruler |
| Pencil, Eraser |
| Pencil, Correctionpen |
| Pencil, Ruler |
| Eraser, Correctionpen |
| Eraser, Ruler |
| Correctionpen, Ruler |

| Table 4.4 | |
|---|---|
| Itemsets | sup_count |
| Pen, Pencil | 4 |
| Pen, Eraser | 4 |
| Pen, Correctionpen | 1 |
| Pen, Ruler | 2 |
| Pencil, Eraser | 4 |
| Pencil, Correctionpen | 2 |
| Pencil, Ruler | 2 |
| Eraser, Correctionpen | 0 |
| Eraser, Ruler | 1 |
| Correctionpen, Ruler | 0 |

Now, the set of frequent 2- itemsets, L2, is determined by candidate support count with minimum support count. L2 is shown in Table 4.5 and that is satisfied the minimum support count.

| Table 4.5 | |
|---|---|
| Itemsets | sup_count |
| Pen, Pencil | 4 |
| Pen, Eraser | 4 |
| Pen, Ruler | 2 |
| Pencil, Eraser | 4 |
| Pencil, Correctionpen | 2 |
| Pencil, Ruler | 2 |

L2 with L2 are joined to get the set of candidate 3- itemsets, C3. Based on the Apriori property that all subset of a frequent itemsets must also be frequent .Therefore final pruning result is shown in Table 4.6. And then the itemsets in transaction database are scanned and the support count of each candidate itemsets in C3 is accumulated, as shown in Table 4.7.

| Table 4.6 |
|---|
| Itemsets |
| Pen, Pencil, Eraser |
| Pen, Pencil, Correctionpen |

| Table 4.7 | |
|---|---|
| Itemsets | sup_count |
| Pen, Pencil, Eraser | 2 |
| Pen, Pencil, Correctionpen | 2 |

Now, the set of frequent 3- itemsets, L3, is determined by candidate support count with minimum support count. L3 is shown in Table 4.8 and that is satisfied the minimum support count

| Table 4.8 | |
|---|---|
| Itemsets | sup_count |
| Pen, Pencil, Eraser | 2 |
| Pen, Pencil, Correctionpen | 2 |

## 5. System design

In this section, the detail of the system is showed with some motivation examples. This system intends to extract the most frequent itemsets from the transaction database.

In this system, the sales Boucher to customer are inputted as data. Such data are stored as transaction data in the database. First of all, user must input minimum support count for finding frequent itemsets. The algorithm scans all of the transactions in order to count the number of occurrences of each item. Each item is a member of the set of candidate one itemsets (C1). Then the itemsets which satisfy the minimum support count are described as one frequent itemsets (L1). Next we join these L1 to get candidate two itemsets (C2).Then the transactions in database are scanned and the count of each candidate itemsets in C2 is accumulated. Next the itemsets which satisfy the minimum support count are described as two frequent itemsets (L2). By joining L2 to L2 we generate candidate three itemsets (C3). By scanning transaction in database, the count of C3 itemsets are shown. Then the three frequent itemsets (L3) which satisfy the minimum support count are generated. Then association rule can be generated from L3. For knowing this rule is strong or not, user must input minimum confidence level. The confidence level of the itemsets which satisfy the minimum confidence level can be shown as strong association rule.

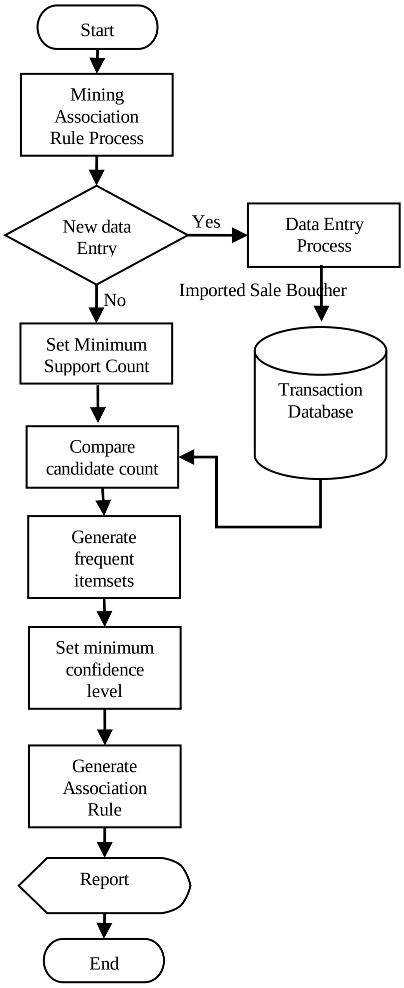


**Figure 1. System Flow Diagram**

## 6. Experimental result

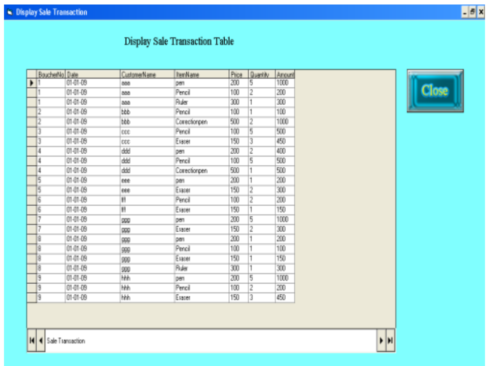


**Figure 2.** Transaction Data

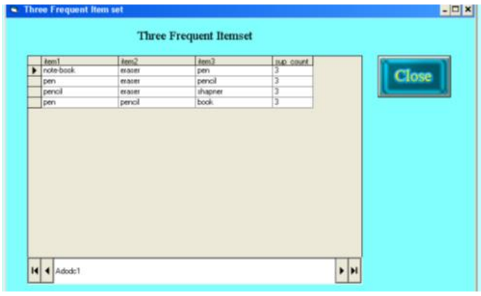


**Figure 3.** Three frequent itemsets list

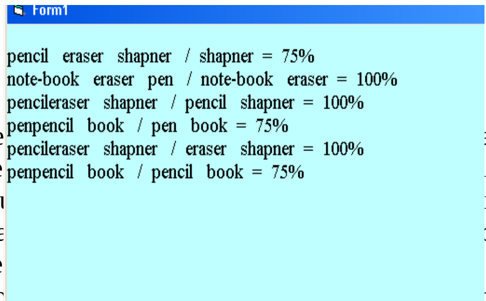


**Figure 4.** Strong association rule

In ... are pre... a in the ... des Bou... ice, qua... ows the ... 6.3 sho... that satisfies the minimum support count 3. Figure-6.4 shows the association rule and Figure 6.5 shows the strong association rule that satisfy the minimum confidence level 70%.

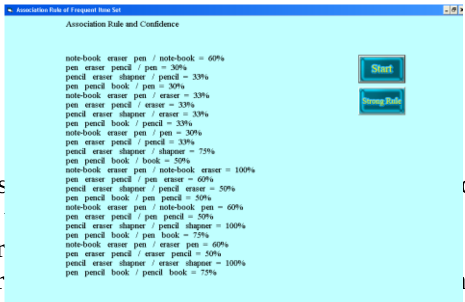


**Figure 5.** Association Rule

By us can use or adver as differ low the best selling itemsets. So the sale manager can allocated the counter of these frequent itemsets in nearly.

## 7. Conclusion

In this paper, a system for mining association rule has been developed and it is based on the data mining system. And then it had been used Apriori Algorithm. Our system is very simple yet powerful because it can obtain more useful and meaningful results, such as frequent itemsets and strong association rule. It can support of mining knowledge and concepts extraction from the transaction data. This system will focus only upon the minimum support count and confidence level. To produce association rules or strong rules, the user must define the inputs, such as, transaction data, support count for frequent itemsets and confidence level for association rule.

## References


[1] D.A.Keim *"Knowledge Discovery and Data Mining"*.

[2] Edward R.Omiecinski *"Alternative interest measures for mining associations in databases"*. IEEE Transactions on Knowledge and Data Engineering, 15(1): 57-69, Jan/Feb 2003.

[3] J. Han and M. Kamber *"Data Mining concept and Techniques"* Morgan Kaufmann, 2001.

[4] Marcos M. Compas *"What is data mining"*, January, 2006.

[5] M. J. Zarki. (2001). SPAD *"An Efficient Algorithm for Mining Frequent Sequences"*. Machine Learning Journal, 42, 31-60.

[6] Pang-Nang, T.M. Steinbanch and V. Kumar *"Introduction to Data Mining"*.

[7] Piatetsky-Shapiro, G. (1991), W.J. Frawely. *"Discovery, Analysis and Presentation of strong rules and Knowledge Discovery in Database,"* AAAI/ MTI Press, Cambridge, MA.

[8] Rakesh Arrival and Ramakrishan Srikand."*Fast Algorithm for Mining Association Rules in Large Database*". Proceeding of the 20th international conference on Very Large Data Bases, VlDB, pages 487-499. Santiago, Chile, September 1994.

[9] R. Aerial; T.Imielinski; A. Swami. *"Mining Association Rules between Sets of Items in Large Databases"*, SIGMOD Conference, 1993: 207-216.

[10] Sergey Brin, Rajeev Motwani, Jeffrey D.Ullman, and Shalom Tsur. *"Dynamic itemset counting and implication rules for market basket data"*.

[11] http:// en.wikipedia. org/ wiki/ Association _ rule_learning.

[12] http:// eruditionhome.com/ dataming.

[13] www.crim.co.umist.ac.uk