INTERNATIONAL JOURNAL OF ADVANCE RESEARCH AND DEVELOPMENT

IJARnD

(Volume 4, Issue 7)
Available online at: www.ijarnd.com

ECLAT based market basket analysis for electronic showroom

Moe Moe Hlaing
Student, Universities of Computer Studies, Magway, Myanmar

ABSTRACT

Market basket analysis is a data mining technique to discover associations between datasets. Association rule mining identifies a relationship between a large set of data items. When a large quantity of data is constantly obtained and stored in databases, several industries are becoming concerned about mining association rules from their databases. Market basket analysis examines customer buying patterns by identifying associations among various items that customers place in their shopping baskets. It is helpful to examine customer purchasing behavior and assists in increasing sales. So, this system is intended to develop a system for market basket analysis on Electronic showroom which will generate association rules among itemsets with the use of ECLAT (Equivalence Class Transformation) algorithm. This system supports the decision-making process for a market expert.

Keywords— *Itemsets, Association Rule, ECLAT*

1. INTRODUCTION

Market basket analysis (association analysis) is a mathematical modeling technique if the customer buys a certain group of items, the customer is likely to buy another group of items. It is used to analyze the customer purchasing behavior, help in increasing the sales and maintain inventory by focusing on the point of sale transaction data. Market basket analysis technique focuses on discovering the purchasing patterns by extracting associations, or co-occurrences from showroom transactional data. It determines the products which are bought together and to recognize the Electronic showroom layout, and also to design promotional campaigns such that products' purchase can be improved. This analysis takes business decisions like what to place on sale, the way to place things on shelves to maximize profit, etc.

Association rule mining is used in customer market analysis. Association rule has recently received much attention from the database community. The process of finding association rules has two separate phases. In the first phase, find all combinations of items that have transaction support above the minimum support count. In the second phase, use the frequent itemsets to generate the desired rules. Most of the previous algorithms are based on the traditional horizontal database format for mining. In a vertical database, each item is associated with its corresponding transaction id (TIDset). Mining algorithms using the vertical format have shown to be very effective and usually outperform horizontal approaches because frequent itemsets can be countered via TIDset intersections in the vertical approach.

So, this system is intended to mine the frequent itemsets on the transaction data of Electronic showroom by using ECLAT algorithm. This algorithm is based on the vertical database format. By using this system, the sale of Electronic showroom is improved in the market. This system can also help the retailers, buyers, planners, merchandisers, and store managers to plan more profitable advertising and promotions, attract more customers and increase the value of the market basket.

2. RELATED WORK

In 2017, G. S. Gayathri [1] compared the performance between Apriori, ECLAT and FP-Growth algorithms for association rule learning. The performance of these three algorithms is compared based on Time efficiency. After the comparison, they present that Apriori algorithm is the fastest algorithm for large dataset and FP-Growth is the fastest algorithm for small dataset. ECLAT algorithm takes less time to generate frequent item sets as compared to other algorithms.

In 2017, P. D. Poonam and M. Emmanuel [2] presented the web pattern mining by using ECLAT algorithm. The web results more relevant to the user query through keyword expansion have been retrieved by the system. This data is being used for the efficient association rule mining using ECLAT algorithm which is weaved for the vertical transactions-based scheme. The results of the ECLAT identify interesting patterns of the extracted web data for the effectiveness.

In 2018, R. Jason and B. T. David [3] used the ECLAT algorithm that is applied to the open-source library SPMF. In this system, this application can perform data mining assisted by open-source SPMF with determined writing format of transaction data. It successfully displayed data with 100% success rate. The application can generate a new easier knowledge which can be used for marketing the product.

3. MARKET BASKET ANALYSIS

Market basket analysis is an effective tool in the retail industry which will help the market owner to increase the business and improve sales distribution techniques. This is totally done by association rule mining in which it analyses the customer behavior against the purchasing item from the market. It analyses the customer purchasing pattern and generates frequent itemsets. After the generation of frequent itemset, it is easy to find the most popular itemset and worst item combination from large transactional data instead of reading it manually. Generation of frequent itemset will enhance the market strategy, placement of goods and many more [5].

Market basket analysis explains the combinations of product that frequently co-occur in transactions. For example, people who buy bread and eggs, also tend to buy butter as many of them are planning to make an omelet. The marketing team should target customers who buy bread and eggs with offers on butter, to encourage them to spend more on their shopping basket. It is also known as “Affinity Analysis” or “Association Rule Mining” [4].

3.1 Association Rule Mining

Association rule mining is one of the most researched areas of data mining and has recently received much attention from the database community. It provides a platform where frequent patterns, correlation, relationship, and associations can be discovered from transactional data kept in various kinds of databases and other data repositories. The aim of association rule within a set of the transaction is to find the rules that can help in making predictions on the occurrence of particular items based on the occurrence of others in the given set of transaction. The original purpose of association rule mining as stated in was to discover interesting frequent patterns, association, correlation or relationships among a large data set items in a given transaction. These can be achieved through the use of support and confidence which is the best-known measures in the evaluation of association rule interestingness [6].

3.2 Benefits of using Association Rule

Association rule might be helpful in designing promotions and discounts or shelf organization and store layout. However, association rules have many other fields in which it has been helpful. Association rules mining is used in the telecommunications and medical fields for performing partial classification. This type of mining has been also used on other types of data sets. It has been used to mine webservers log files to discover the patterns that access different resources consistently and occur together or the access of a particular place occurring at regular times [6].

4. ECLAT ASSOCIATION RULE MINING

ECLAT (Equivalence Class Transformation) association rule mining method generates frequent items only once. Frequent itemsets are those items which frequently occur in the database. There is a number of algorithms for finding frequent itemsets. Apriori is a basic algorithm for finding frequent itemsets. But it takes more time for finding the frequent itemsets, It needs to scan the database again and again which is a time-consuming process. ECLAT algorithm is developed to remove the limitations of Apriori, algorithm.

ECLAT algorithm uses a vertical database. By which it needs to scan the database only once. First, transform the horizontally formatted data to the vertical format by scanning the data set once. Mining can be performed on this data set by intersecting the TID-sets of every pair of a frequent single item. The support count of an itemset is simply the length of the TID-set of the itemset. If the minimum support count is 2, the association rules can be generated from any frequent itemsets. ECLAT employs an optimization called “fast intersection,” in that whenever two TID-lists are intersected, we only consider the resulting TID-list if its cardinality reaches minimum support. In other words, each intersection is eliminated as soon as it does not meet the minimum support [7].

4.1 ECLAT Algorithm

ECLAT algorithm is as follows:

```
Input: D, s, I ⊆ I
Output: F [I] (D, s)
F [I]: = {}
for all i ∈ I occurring in D do
F [I]: =  F [I] U {I U {i}}
 //Create D^i
D^i : = {}
for all j  ∈ I occurring in D such that j>i do
C: = cover ({i})∩cover ({j})
if|C|≥ s then
D^i : = D^i U {(j, C)}
end if
end for
//Depth-first recursion
Compute   F [I U  {i}]( D^i, s)
F [I]: = F [I] U F [I U {i}]
end for
```

4.2 Utility Function

The potential usefulness of a pattern is a factor defining its interestingness. It can be estimated by a utility function, such as support. Rule A⇨B (A and B are set of items) has support s if s% of all transaction contains both A and B [8].

$$\text{Support}(\text{"A} \rightarrow \text{B"}) = \frac{\#\text{tuples\_Containing\_both AandB}}{\text{total\_\#\_oftuples}}$$

4.3 Certainty Function

A certainty measure for association rules of the form “A⇨B”, where A and B are sets of item sets is confidence. Rule A⇨B (A and B are set of items) has confidence c if c% of transactions that contains A also contain B [8].

$$\text{Confidence}(\text{"A} \rightarrow \text{B"}) = \frac{\#\text{tuples\_Containing\_both AandB}}{\#\text{tuples\_Containing\_A}}$$

5. PROPOSED SYSTEM DESSIGN

The proposed system is implemented to find out which items are commonly purchased together within the electronic showroom in order to make some selected frequent customers special bundle-offers which are likely to be in their interest. This system searches the interesting relationships among items by using ECLAT algorithm. These are step by step processing to generate an association rule. Firstly, this system extracts the transaction from an electronic showroom transactional database. Then, it is compared with the minimum support value. Items less than minimum support value is removed and others go on processing. And then, this system again compares each of them with minimum support value and removes pairs which are less than minimum support value.

After finishing this processing, this system produces an association rule which is generated by using ECLAT. The rules having equal to or greater confidence than user-specified one is considered to be a strong association rule. And then, this system measures the processing time as the performance of ECLAT. Finally, this system displays the processing time and association result.

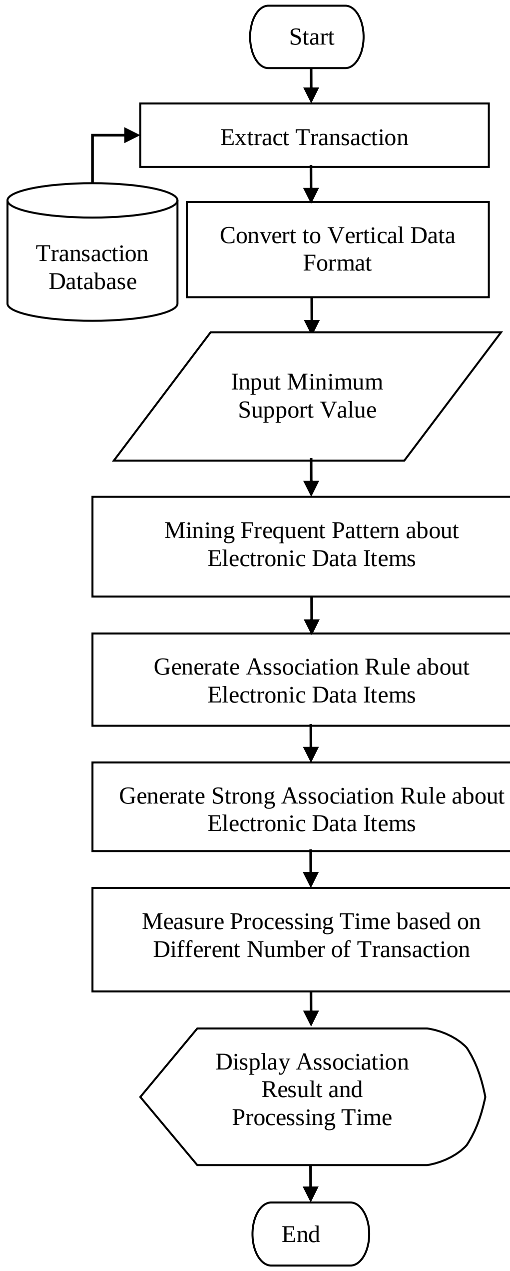

Fig. 1: Proposed system design

5.1 Explanation of the System

As a sample, this system tested seven transactions with electronic data in the showroom. In this sample explanation, this system uses the "3" minimum support value (count). Minimum support count can be defined as 2 and more than 2. Sample transaction database is shown in Table 1.

Table 1: Transaction Database

TID	Items
1	TV, EVD, Extension-Cord
2	Iron, Extension-Cord, Rice-cooker, Battery-charger, Pressure-cooker
3	Laptop, Rice-cooker, Battery-charger, Pressure-cooker
4	Memory-stick, TV, EVD, Iron, Extension-Cord
5	Pressure-cooker, Woofer, Battery-charger, Extension-Cord
6	Iron, Extension-Cord, Rice-cooker, Battery-charger, Pressure-cooker
7	EVD, TV, Battery-charger, Pressure-cooker

Initially, every item which is equal to and greater than the user-specified minimum support count is considered as a frequent 1-itemset from the transaction. Frequent 1-itemset is shown in Table 2.

Table 2: Frequent 1-Itemset

Itemset	TID_Set	Support Count
{TV}	{1,4,7}	3
{EVD}	{1,4,7}	3
{Extension-Cord}	{1,2,4,5,6}	5
{Iron}	{2,4,6}	3
{Rice-cooker}	{2,3,6}	3
{Battery-charger}	{2,3,5,6,7}	5
{Pressure-cooker}	{2,3,5,6,7}	5

In the next iteration, frequent 2-itemsets are generated. Frequent 2-itemset is shown in Table 3.

Table 3: Frequent 2-Itemset

Itemset	TID_Set	Support Count
{TV, EVD}	{1,4,7}	3
{Extension-Cord, Iron}	{2,4,6}	3
{Extension-Cord, Battery-charger}	{2,5,6}	3
{Extension-Cord, Pressure-cooker}	{2,5,6}	3
{Rice-cooker, Battery-charger}	{2,3,6}	3
{Rice-cooker, Pressure-cooker}	{2,3,6}	3

There are no frequent itemsets at this process, iterated procedures are terminated. Frequent 3-itemset is shown in Table 4.

Table 4: Frequent 3-Itemset

Itemset	TID_Set	Support Count
{Rice-cooker, Battery-charger, Pressure-cooker}	{2, 3, 6}	3
{Extension-Cord, Battery-charger, Pressure-cooker}	{2, 5, 6}	3

Thus, the frequent 3-itemsets {Rice-cooker, Battery-charger, Pressure-cooker} and {Mouse, Battery-charger, Pressure-cooker} have been received with twelve rules. Association rules are shown in Table 5.

Table 5: Association Rules

Subset(A)	Subset(B)	Confidence (%)
{Rice-cooker}	{Battery-charger, Pressure-cooker}	100
{Battery-charger}	{Rice-cooker, Pressure-cooker}	60
{Pressure-cooker}	{Rice-cooker, Battery-charger}	60
{Rice-cooker, Battery-charger}	{Pressure-cooker}	100
{Battery-charger, Pressure-cooker}	{Rice-cooker}	75
{Rice-cooker, Pressure-cooker}	{Battery-charger}	100
{Extension-Cord}	{Battery-charger, Pressure-cooker}	60
{Battery-charger}	{Extension-Cord, Pressure-cooker}	60
{Pressure-cooker}	{Extension-Cord, Battery-charger}	60
{Extension-Cord, Battery-charger}	{Pressure-cooker}	100

{Extension-Cord, Pressure-cooker}	{Battery-charger}	100
{Battery-charger, Pressure-cooker}	{Extension-Cord}	60

In this system, each user is allowed to enter minimum confidence to produce strong rule. Since, the rules, having equal to or greater confidence than the user-specified one, are considered to be strong. In this example, the minimum confidence is 100%. Strong association rules are as follows:

- Rule 1: {Rice-cooker} ► {Battery-charger, Pressure-cooker}
- Rule2: {Rice-cooker, Battery-charger} ► {Pressure-cooker}
- Rule 3: {Rice-cooker, Pressure-cooker} ► {Battery-charger}
- Rule 4: {Extension-Cord, Battery-charger} ► {Pressure-cooker}
- Rule 5: {Extension-Cord, Pressure-cooker} ► {Battery-charger}

5.2 Experimental result of the system

By using association rule mining, this system analyzes the itemsets pairs that likely to happen for future sales transactions. This system tested the 200 transactions to measure the performance of the system. ECLAT algorithm does not need to scan the database to find support. So, the system that uses the ECLAT algorithm can take less processing time for association rule production. Processing time result is shown in Table 6.

Table 6: Processing Time Results

ID	Number of Transactions	Processing Time (milliseconds)
1	50	15
2	100	18
3	150	26
4	200	30

The performance of the system is shown in figure 2.

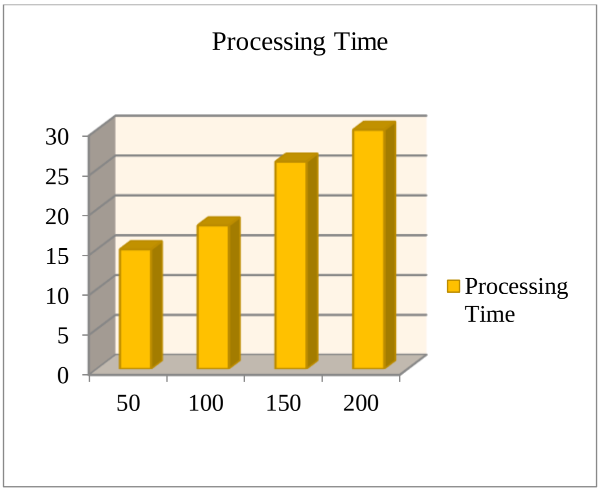

Fig. 2: Performance of the System

6. CONCLUSION

In this system, association rule mining is implemented on the basis of the ECLAT algorithm. By taking less processing time, this algorithm is fast to produce the association rules. Therefore, this system provides the decision-maker to give useful information about interesting items. This system is also a provider of several devices and business organizations. The system is implemented by collecting real data from Electronic Showroom. So, this system can also support this electronic showroom manager who can place the related devices together and advice the customer for the best price and the latest updates.

7. REFERENCES

[1] G. S. Gayathri, "Performance Comparison of Apriori, ECLAT and FP-Growth Algorithm for Association Rule Learning", International Journal of Computer Science and Mobile Computing (IJCSMC), vol. 6, no. 2, pp. 81-89, 2017.

[2] P. D. Poonam and M. Emmanuel, "Web Pattern Mining using ECLAT", International Journal of Computer Applications, vol. 179, no. 8, pp. 9-14, 2017.

[3] R. Jason and B. T. David, "Data Mining Application using Association Rule Mining ECLAT Algorithm based on SPMF", MATEC Web of Conferences, pp. 1-10, 2018.

[4] R. Karthiyayini and R.Balasubramanian, "Affinity Analysis and Association Rule Mining using Apriori Algorithm in Market Basket Analysis", International Journal of Advanced Research in Computer Science and Software Engineering, vol. 6, no. 10, pp. 241-246, 2016.

[5] B. Mahesh, F. Ankush and B. Aniket, "Market Basket Analysis based on Frequent Itemset Mining", International Research Journal of Engineering and Technology (IRJET), vol. 5, no. 2, pp. 414-417, 2018.

[6] A. A. Zhang, N. Goga and S. Kuyoro, "An Overview of Association Rule Mining (ARM) Algorithms for Market Basket Analysis (MBA)", Journal of Research in Engineering and Applied Sciences, vol. 2, no. 4, 132-140, 2017.

[7] K. Vani, "Comparative Analysis of Association Rule Mining Algorithms based on Performance Survey", International Journal of Computer Science and Information Technologies, vol. 6, no. 4, pp. 3980-3985, 2015.

[8] H. Jiawei, K. Micheline, "Data Mining: Concepts and Techniques", Simon Fraser University, US, 2001.