MARKET BASKET ANALYSIS USING FP GROWTH AND APRIORI ALGORITHM: A CASE STUDY OF MUMBAI RETAIL STORE

Prof. Kavitha Venkatachari (kavitav@ibsindia.org),
Issac Davanbu Chandrasekaran IBS Business School, Powai, Mumbai, India

Abstract

Companies nowadays are rich in vast amounts of data but poor in information extracted from that data. Big data is seen as a valuable resource and although the concept of data mining is still new and developing, companies in a variety of industries are relying on it for making strategic decisions. Facts that otherwise may go unnoticed can be now revealed by the techniques that sift through stored information. Market basket analysis is a very useful technique for finding out co-occurring items in consumer shopping baskets. Such information can be used as a basis for decisions about marketing activity such as promotional support, inventory control and cross-sale campaigns. The main objective of the research paper is to see how different products in a grocery store assortment interrelate and how to exploit these relations by marketing activities. Mining association rules from transactional data will provide us with valuable information about co-occurrences and co-purchases of products. Such information can be used as a basis for decisions about marketing activity such as promotional support, inventory control and cross-sale campaigns. To find association rules we use two algorithms.(FP growth and Apriori algorithm).To find out frequent item sets using R tool and Rapid miner. As per the paper FP growth is much slow in Rapid Miner and in R Programming Apriori algorithm is fast. We have collected data from Mumbai Retail Store and the sample size for the analysis is 300.

Keywords: Market Basket Analysis, FP Growth and Apriori algorithm, cross-sale campaigns, promotional support, inventory control, and frequent item sets, data mining, association.

Business use of data mining

Data mining is commonly seen as a single step of a whole process called Knowledge Discovery in Databases (KDD). According to Fayyad et.al, 'KDD is the nontrivial process of identifying valid, novel, potentially useful and ultimately understandable patterns in data.' (Usama Fayyad, Gregory Piatetsky-Shapiro, and Padhraic Smyth, 1996)

Data mining is a technique that encompasses a huge variety of statistical and computational techniques such as: association-rule mining, neural network analysis, clustering, classification, summarizing data and of course the traditional regression analyses.

Objective of the study

In the recent years analyzing shopping baskets has become quite appealing to retailers. Advanced technology made it possible for them to gather information on their customers and what they buy. The introduction of electronic point-in sale increased the use and application of transactional data in market basket analysis. In retail business analyzing such information is highly useful for understanding buying behaviour. Mining purchasing patterns allows retailers to adjust promotions, store settings and serve customers better.

Identifying buying rules is crucial for every successful business. Transactional data is used for mining useful information on co-purchases and adjusting promotion and advertising accordingly. The well-known set of beer and diapers is just an example of an association rule found by data scientists.

The main objective of this research is to see how different products in a retail shop assortment interrelate and how to exploit these relations by marketing activities. Mining association rules from transactional data will provide us with valuable information about co-occurrences and co-purchases of products. Some shoppers may purchase a single product during a shopping trip, out of curiosity or boredom, while others buy more than one product for efficiency reasons.

Motive of the Research

It is very important for retailers to get to know what their customers are buying. Some products have higher affinity to be sold together and hence the retailer can benefit from this affinity if special offers and promotions are developed for these products. It is also important to the retailer to cut off products from the assortment which are not generating profits. Deleting loss-making, declining and weak brands may help companies boost their profits and redistribute costs towards aspects of the more profitable brands. (Kumar, 2009). This is yet another reason why data mining is seen as a powerful tool for many businesses to regularly check if they are selling too many brands, identify weak ones and possibly merge them with healthy brands.

Data mining techniques are highly valued for the useful information they provide so that the retailer can serve customers better and generate higher profits.

1. Find products with affinity to be sold together.
2. Improve in-store settings and optimize product placement.
3. Improve layout of the catalogue of e-commerce site.
4. Control inventory based on product demand.

Literature Review

Data mining has taken an important part of marketing literature for the last several decades. Market basket analysis is one of the oldest areas in the field of data mining and is the best example for mining association rules.

Various algorithms for Association Rule Mining (ARM) and Clustering have been developed by researchers to help users achieve their objectives. Rakesh Agrawal and Usama Fayyad are one of the pioneers in data mining. They account for a number of developed algorithms and procedures.

According to Shapiro, rule generating procedures can be divided into procedures that find quantitative rules and procedures that find qualitative rules. (Rakesh Agrawal, Ramakrishnan Srikant) elaborate on the concept of mining quantitative rules in large relational tables. Quantitative rules are defined in terms of the type of attributes contained in these relational tables. Attributes can be either quantitative (age, income, etc.) or categorical (certain type of a product, make of a car). Boolean attributes are such attributes that can take on one of two options (True or False, 1 or 0). They are considered a special case of categorical attributes. The authors call this mining problem the Quantitative Association Rules problem. An example of a generated quantitative rule is :

If ((Age : [30…39]) + (Married : Yes)) → (Number of cars = 2)

The example combines variables that have quantitative and boolean attributes.

(S. Prakash, R.M.S. Parvathi, 2011) propose a qualitative approach for mining quantitative association rules. The nature of the proposed approach is qualitative because the method converts numerical attributes to binary attributes.

However, finding qualitative rules is of main interest in this analysis. These rules are most commonly represented as decision trees, patterns or dependency tables. (Gregory Piatetsky-Shapiro, William Frawley, 1991) The type of attributes used for mining qualitative rules is categorical.

(Rakesh Agrawal, Tomasz Imielinski, Arun Swami, 1993) is one of the first published papers on association rules that proposes a rule mining algorithm that discovers qualitative rules with no restriction for boolean attributes. The authors test the effectiveness of the algorithm by applying it to data obtained from a large retailing company.

Association rules found application in many research areas such as: market basket analysis, recommendation systems, intrusion detection etc.

In marketing literature market basket analysis has been classified into two models: explanatory and exploratory. First, exploratory models will be thoroughly explained in this paper as they are of higher relevance for the research and after that an explanation of explanatory models will be given. The main idea behind exploratory models is the discovering of purchase patterns from POS (point-of-sale) data. Exploratory approaches do not include information on consumer demographics or marketing mix variables. (KatrinDippold, Harald Hruschka, 2010) Methods like association rules (Rakesh Agrawal, Sirkant Ramakrishnan, 1994) or collaborative filtering (Andreas Mild, Thomas Reutterer, 2003) summarise a vast amount of data into a fewer meaningful rules or measures. Such methods are quite useful for discovering unknown relationships between the items in the data. Moreover, these methods are computationally simple and can be used for undirected data mining. However, exploratory approaches are not appropriate for forecasting and finding the cause-roots of complex problems. They are just used to uncover distinguished cross-category interdependencies based on some frequency patterns for items or product categories purchased together.

A typical application of these exploratory approaches is identifying product category relationships by simple association measures. Pair wise associations are used to compare entities in pairs and judge which entity is preferred or has greater amount of some quantitative property. (Julander, 1992) compares the percentage of shoppers buying a certain product and the percentage of all total sales generated by this product. By making such comparisons, one can easily find out the leading products and what is their share of sales. Examining which the leading products are for consumers is extremely important since a large number of shoppers come into contact with these specific product types every day.

As the departments with leading products generate much in-store traffic, it is crucial to use this information for placing other specific products nearby. The paper by Julander also shows how combinatory analysis can be used to study the patterns of cross-buying between certain brands or product groups: for instance, what is the percentage of shoppers that buy products A+C, but not B or what is the percentage of shoppers that buy only A. It also deals with the probabilities that shoppers will purchase from one, two or more departments in a single visit in the store.

Another significant stream of research in the field of exploratory analysis is the process of generating association rules. Substantial amount of algorithms for mining patterns from market basket data have been proposed. From the co-operative work of Rakesh Agrawal and Ramakrishnan Srikant they present two new algorithms for discovering large itemsets in databases, namely Apriori and AprioriTid. These two algorithms are similar with regard to the function that is used to determine the candidate item sets, but the difference is that the AprioriTID does not use the database for counting support after the first pass (first iteration) while Apriori makes multiple passes over the database (more information on methodology in Chapter 4).

The results from the study show that these two new algorithms perform much better than the previously known AIS (R. Agrawal, T. Imielinski, and A. Swami, 1993) and SETM (M. Houtsma and A. Swami, 1993) algorithms. Since the introduction of the Apriori algorithm, it has been considered the most useful and fast algorithm for finding frequent item sets.

Many improvements have been made on the Apriori algorithm in order to increase its efficiency and effectiveness. (M.J.Zaki, M.Ogihara, S. Parthasarathy, 1996). There are few algorithms developed that are not based on the Apriori,but they still address the issue of speed of Apriori. The following papers (Eu-Hong (Sam) Han, George Karypis, Vipin Kumar, 1999) , (Jong Soo Park, Ming-Syan Chen, Philip S. Yu) propose new algorithms which are not based on the Apriori, but all of them are being compared to Apriori in terms of execution time.

Data Pre-processing

Market Basket Transactions

The transaction data set can be used to keep events in the form of a file where each record represents the transaction. This section presents a methodology known as market basket transactions, which is useful for discovering interesting relationships hidden in big datasets. Table 1 illustrates an example of market basket transactions.

Table 1 Example of market basket transaction

ID	Item Sets
1	{ Bread, Butter}
2	{ Jam, Dairy, Canned Foods}
3	{ Jam, Breads, Butter, Paper_goods}
4	{Canned Foods, Dairy, Bread, Butter}

Consider in Table 1, the following rule can be extracted from the database is shown in Figure 1. {Bread}→{Butter}

(A)

ID	Bread	Butter	Jam	Juices	Paper_goods	Dairy	Canned
							Foods
1	1	1	0	1	0	1	1
2	0	0	1	1	1	0	1
3	1	0	1	0	1	0	1
4	1	1	1	0	1	0	0
5	0	1	0	1	1	1	0

(B)

	Bread	Butter	Jam	Juices	Paper_goods	Dairy	Canned
							Foods
Item set	1	1	2	1	2	1	1
	3	4	3	2	3	5	2
	4	5	4	5	4		3
					5		

(C)

ID	Item Set
	Bread,Butter,Juices, Dairy products, Canned Foods
2	Jam,Juices,Paper Goods,Canned Goods
	Bread,Jam,Paper Goods,Canned goods
4	Bread,Butter,Jam,Paper Goods
5	Butter,Juices,paper Goods,Dairy products

The rule suggests that a strong relationship because many customers who by breads also buy butter. Retailers can use this type of rules to them identify new opportunities for cross selling their products to the customers.

Binary Representation

Market basket data can be represented in a binary format as shown in Figure 1, where each row corresponds to a transaction and each column corresponds to an item. This format is usually used when association rule mining is used to identify frequency of item sets.

Figure 1 An example supermarket database with five transactions (A) Binary database, (B) Vertical Database, (C) Transaction Database.

Association rule algorithm

Let I = {i1, i2,…, in} be set of items (n binary attributes). Let D = {t1, t2,…,tn} be a database having a set of transactions where each transaction in D has an unique transaction T and contains a subset of the items in I. An association rule is association relationship of the form x□y, where x Ì i, y Ì i, and x Ç y = Æ .

The support of rule X and Y are called antecedent (LHS: left hand side) and consequent (RHS: right hand side) of the rule. Consider in Figure 1, the set of items is I = {Breads, Butter, Juices, Dairy, Canned Foods}. An example rule for the supermarket could be {Breads, Butter}□{Juices} meaning that if Breads and Butter is bought, customers also buy Juices. The strength of an association rule can be measured in terms of its support and confidence. Support determines how often a rule is applicable to a given data set, while confidence determines how frequently items in Y appear in transactions that contain X. The formal definitions of support, confidence, and lift are defined in Eq. (1), Eq. (2), and Eq. (3), respectively.

$$\text{Support}(X \rightarrow Y) = \frac{\text{P robability}(X \cup Y)}{\text{Total number of transactions}} \quad (1)$$

$$\text{Confidence}(X \rightarrow Y) = \frac{\text{P robability}(X \cap Y)}{\text{Number of t ransact ion}(X)} \quad (2)$$

$$\text{Lift } (X \rightarrow Y) = \frac{\text{P robability}(X \cap Y)}{\text{P robability}(X)\text{P robability}(Y)} \quad (3)$$

Proof of Equation

Consider from Eq. (1 and 2), if the rule {Jam, Bread}→{Butter}. Therefore, the support count for {Jam, Bread, Butter} is 3 and the total number of transactions is 5, the rule's support is 3/5 = 0.6. While the rule's confidence is obtained by dividing the support count for {Jam, Bread, Butter} by the support count for {Bread, Butter}. Since there 3 transactions that contain Breads and Butter, the confidence for this rule is 2/3 = 0.67.

Frequent Item Set Generation by using FP-Growth Algorithm

The FP-growth is algorithm which overcomes the major problems association rule with Apriori algorithm. It follows a divide and conquer strategy. The original dataset is transformed into a tree well known as FP-tree, which holds all the information regarding frequent items. The transformed dataset is divided into a set of conditional dataset for each frequent item and mines each such dataset separately to generate frequent items. In this work, we have used this algorithm for generating frequent item sets mining from market basket analysis. All frequent item sets can be mined from the tree directly via the FP-growth algorithm, whose pseudo-code is shown in Figure 2, which proposed by [9].

FP-Growth Algorithm

Figure 2 Pseudo-code of FP-growth algorithm

```
// Initial Call: R←FP-tree(D), P← Æ , F← Æ
FP-growth (R, P, F, minsup):
Remove infrequent items from R
if IsPATH (R) then
foreach Y Í R do
X←P È Y
Sup(X) ←minxєY{cnt(x)}
F←FU{(X, sup(X))}
else
foreach i Î R in increasing order of sup(i) do
X←PU{i}
sup(X) ←sup(i)
F←FU {(X, sup(X))}
RX← Æ //projected FP-tree for X
foreach path Î PATH FROM ROOT(i) do
cnt(i) ← count of i in path
Insert path, excluding I, into FP-tree RX with count
cnt(i)
if Rx ≠ Æ  then FP-growth (Rx, X, F, minsup)
```

Consider from FP-Growth algorithm, the root of the FP-tree is labelled with"null" in the first stage. Afterward, each transaction from the structure are processed in reverse order and saved the number of transactions in the FP-tree structure in reverse order because the aim is to have a rather small tree size, the most frequent articles within the transactions being saved as close as possible to the root.

Developing a Series of Pre-processing for Generating Associations Rule

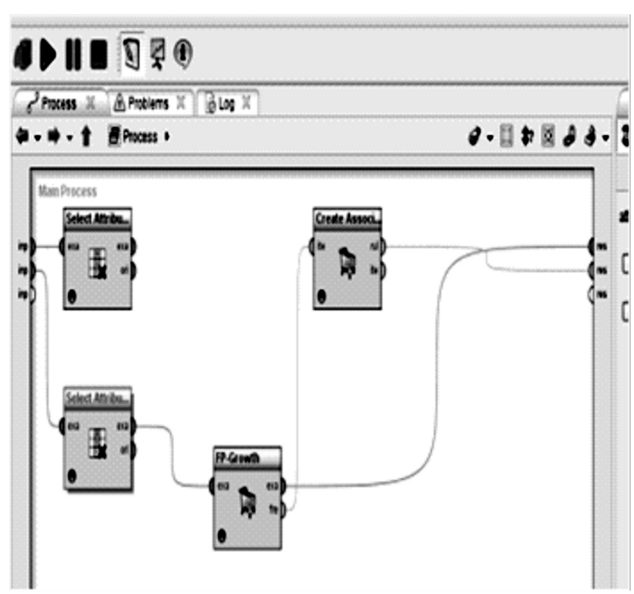
The first stage, to determine the frequent sets and to generate association rules based on the frequent sets discovered. The generating associations rule is presented in Figure 3 through Rapid Miner.

The FP-growth algorithm to determine the frequent item sets and the create association rules algorithm to generate association rules based on the frequent item sets discovered. This algorithm calculates all frequent item sets, building a FP-tree structure from a database transactions and all process are descries below.

1. Retrieve Transaction

This operator can be used to access the repositories. It should replace all file access, since it provides full Meta data processing. In contrast to accessing a raw file, it provides the complete of the data, so all data transformations are possible.

Figure 3 Scheme of the generating association rules by using the FP-growth algorithm (Rapid Miner)



2. Numerical to Binominal

This operator changes the type of the selected numeric attributes to a binominal type. It also maps all values of these attributes to corresponding binominal values. Binominal attributes can have only two possible values i.e. “1” or “0”. If the value of an attribute is between the specified minimal and maximal value, it becomes “0”, otherwise “1”. Minimal and maximal values can be specified by the min and max parameters respectively. If the value is missing, the new value will be missing.

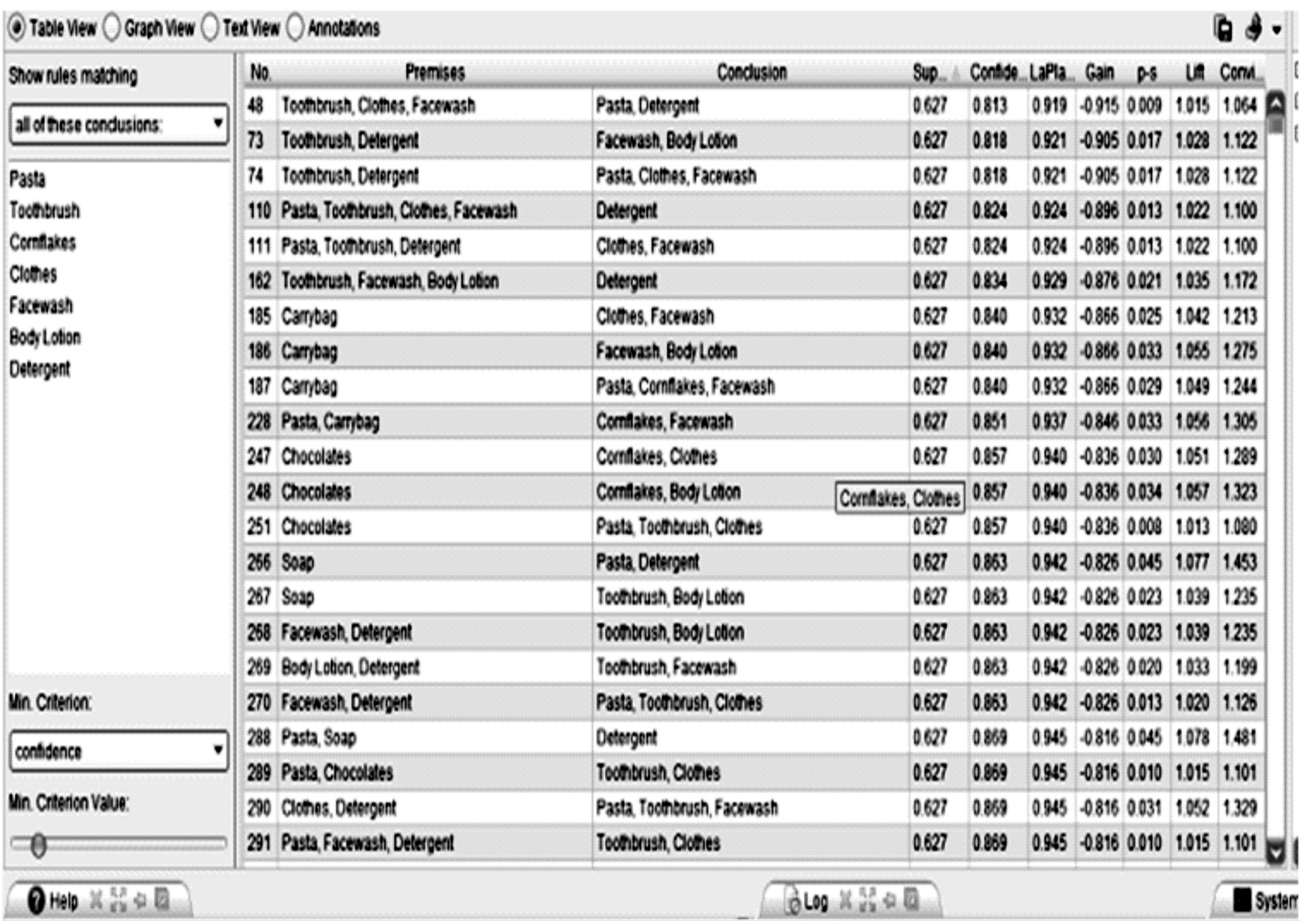
3. FP-Growth Algorithm

This operator calculates all frequent item sets from an example set by building a FP-tree data structure on the transaction data base. This is a very compressed copy of the data which in many cases fits into main memory even for large data bases. All frequent item sets are derived from this FP-tree. Many other frequent item set mining algorithms also exist e.g. the Apriori algorithm. A major advantage of FP-growth algorithm compared to Apriori algorithm is that it uses only 2 data scans and is therefore often applicable even on large data sets.

4. Create Association Rules

Association rules are created by analyzing data for frequent if/then patterns and using the criteria support and confidence to identify the most important relationships. Support is an indication of how frequently the items appear in the database. Confidence indicates the number of times the if/then statements have been found to be true. The frequent if/then patterns are mined using the operators like the FP-growth algorithm. The create association rules algorithm takes these frequent item sets and generates association rules.

Experimental output



No.	Premises	Conclusion	Sup...	Confide...	LaPla...	Gain	p-s	Lift	Conv...
48	Toothbrush, Clothes, Facewash	Pasta, Detergent	0.627	0.813	0.919	-0.915	0.009	1.015	1.064
73	Toothbrush, Detergent	Facewash, Body Lotion	0.627	0.818	0.921	-0.905	0.017	1.028	1.122
74	Toothbrush, Detergent	Pasta, Clothes, Facewash	0.627	0.818	0.921	-0.905	0.017	1.028	1.122
110	Pasta, Toothbrush, Clothes, Facewash	Detergent	0.627	0.824	0.924	-0.896	0.013	1.022	1.100
111	Pasta, Toothbrush, Detergent	Clothes, Facewash	0.627	0.824	0.924	-0.896	0.013	1.022	1.100
162	Toothbrush, Facewash, Body Lotion	Detergent	0.627	0.834	0.929	-0.876	0.021	1.035	1.172
185	Carrybag	Clothes, Facewash	0.627	0.840	0.932	-0.866	0.025	1.042	1.213
186	Carrybag	Facewash, Body Lotion	0.627	0.840	0.932	-0.866	0.033	1.055	1.275
187	Carrybag	Pasta, Cornflakes, Facewash	0.627	0.840	0.932	-0.866	0.029	1.049	1.244
228	Pasta, Carrybag	Cornflakes, Facewash	0.627	0.851	0.937	-0.846	0.033	1.056	1.305
247	Chocolates	Cornflakes, Clothes	0.627	0.857	0.940	-0.836	0.030	1.051	1.289
248	Chocolates	Cornflakes, Body Lotion	Cornflakes, Clothes	0.857	0.940	-0.836	0.034	1.057	1.323
251	Chocolates	Pasta, Toothbrush, Clothes	0.627	0.857	0.940	-0.836	0.008	1.013	1.080
266	Soap	Pasta, Detergent	0.627	0.863	0.942	-0.826	0.045	1.077	1.453
267	Soap	Toothbrush, Body Lotion	0.627	0.863	0.942	-0.826	0.023	1.039	1.235
268	Facewash, Detergent	Toothbrush, Body Lotion	0.627	0.863	0.942	-0.826	0.023	1.039	1.235
269	Body Lotion, Detergent	Toothbrush, Facewash	0.627	0.863	0.942	-0.826	0.020	1.033	1.199
270	Facewash, Detergent	Pasta, Toothbrush, Clothes	0.627	0.863	0.942	-0.826	0.013	1.020	1.126
288	Pasta, Soap	Detergent	0.627	0.869	0.945	-0.816	0.045	1.078	1.481
289	Pasta, Chocolates	Toothbrush, Clothes	0.627	0.869	0.945	-0.816	0.010	1.015	1.101
290	Clothes, Detergent	Pasta, Toothbrush, Facewash	0.627	0.869	0.945	-0.816	0.031	1.052	1.329
291	Pasta, Facewash, Detergent	Toothbrush, Clothes	0.627	0.869	0.945	-0.816	0.010	1.015	1.101

FP Growth algorithm Using Rapid Miner Output

Apriori algorithm using R programming code

```
library(xlsx)
library(arulesviz)

G3_300 <- read.xlsx("C:/Users/shantanu/Desktop/G3_300.xlsx",sheetName = "Sheet1")


#data(Groceries)

rules <- apriori(Big_Data, parameter = list(supp = 0.1, conf = 0.8))

# Show the top 10 rules, but only 2 digits
options(digits=2)
rules<-sort(rules, by="confidence", decreasing=TRUE)

inspect(rules[1:7])
```

Apriori algorithm using R programming output

```
> inspect(rules[1:7])
  lhs             rhs         support confidence lift
1 {Chocolates,
   Chips}      => {Biscuits}    0.11       0.95  1.6
2 {Chips,
   Pulses}     => {Biscuits}    0.10       0.95  1.6
3 {Masala,
   Chocolates} => {Biscuits}    0.12       0.85  1.4
4 {Chocolates} => {Biscuits}    0.22       0.85  1.4
5 {Masala,
   Soap}       => {Biscuits}    0.11       0.84  1.4
6 {Chips}      => {Biscuits}    0.24       0.84  1.4
7 {Noodles,
   Soap}       => {Biscuits}    0.10       0.83  1.4
>
```

Conclusion

A synthetic data set has been used with 77 items each for analysis. A set of association rules are obtained by applying Apriori algorithm and FP growth. By analyzing the data,and giving different support and confidence values,we can obtain different number of rules. During analysis we have obtained different number of rules. During analysis it found that FP growth is very slow for large number of transactions as compare to Apriori.It takes more time to generate frequent item sets in Rapid Miner. We work on a synthetic data which contains 300 transactions. All these results are collected from Intel core i5 CPU with speed and 4GB RAM. In R programming Apriori algorithm generate multiple rules in millisec.

References

M. Khattak, A. M. Khan, Sungyoung Lee and Young-Koo Lee. (2010). Analyzing Association Rule Mining and Clustering on Sales Day Data with XLMiner and Weka. International Journal of Database Theory and Application Vol. 3, No. 1.

Anderson, C. (2006). The Long Tail: Why the Future of Business is Selling Less of More.

Andreas Mild, Thomas Reutterer. (2003). An improved collaborative filtering approach for predicting cross-category purchases based on binary market basket data. Journal of Retailing and Consumer Services vol.10, 123-133.

twareSupportAprioriRulesR Programming0.051millisec111 0.0091sec2090141 0.012sec2090141Rapid MinerFP GrowthRules 0.91sec987 0.05out of order 1. Andreas Mild, Thomas Reutterer. (2003). An improved collaborative filtering approachfor predicting cross-category purchases based on binary market basket data. Journal of Retailing and Consumer Services, Volume 10, 123-133.

Software	Support	Apriori	Rules
R Programming	0.05	1millisec	111
	0.009	1sec	2090141
	0.01	2sec	2090141
Rapid Miner		**FP Growth**	**Rules**
	0.9	1sec	987
	0.05	out of order	

Andrew Ainslie, Peter E. Rossi. (1998). Similarities in Choice Behavior Across Product Categories. Marketing Science, Vol. 17, No. 2, 91-106.

Assunc¸ ˜ao, J. L., & Meyer, R. J. (1993). The rational effect on price promotions on sales and consumption. Management Science, 39 (May), 517–535.

Bari A. Harlam and Leonard M. Lodish. (1995). Modeling Consumers' Choices of Multiple Items. Journal of Marketing Research, Vol. 32, No. 4, 404-418.

Bill Merrilees and Dale Miller. (2001). Superstore interactivity: a new self-service paradigm of retail service? International

Journal of Retail & Distribution Management, Vol. 29, Number 8, 379389.

Byung-Do Kim, Kannan Srinivasan, Ronald T. Wilcox. (1999). Identifying Price Sensitive Consumers: The Relative Merits of Demographic vs. Purchase Pattern information. Journal of Retailing, Volume 75(2), 173-193.

Coenen, F. (2011). Data Mining: Past, Present and Future. The Knowledge Engineering Review, Vol. 26:1, 25-29.

D.W. Cheung, A.W. Fu, and J. Han. (1994). Knowledge discovery in databases: a rule based attributeoriented approach. The 8th International Symposium on Methodologies for Intelligent Systems (ISMIS'94),, (pp. 164-173). Charlotte, North Carolina.

David R. Bell and James M. Lattin. (2008). Shopping Behavior and Consumer Preference for Store Price Format: Why "Large Basket". Marketing Science, Vol. 17, No. 1, 66-88.

David R. Bell and YaseminBoztu˘g. (2007). The positive and negative effects of inventory on category purchase: An empirical analysis. Marketing Letters, 18, 1-14.

Eu-Hong (Sam) Han, George Karypis, Vipin Kumar. (1999). Scalable Parallel Data Mining for Association Rules. IEEE Transactions on Knowledge and Data Engineering, vol.20.

Francis J. Mulhern and Robert P. Leone. (1991). Implicit Price Bundling of Retail Products: A Multiproduct Approach to Maximizing Store. Journal of Marketing,Vol. 55, No. 4, 63-76.

Garry J.Russel, Wagner A. Kamakura. (1997). Modeling Multiple Category Brand Preference with Household Basket Data. Journal of Retailing, Volume 73(4), 439-461.

Gary J Russell, Ann Petersen. (2000). Analysis of Cross-Category Dependence in Market Basket Selection. Journal of Retailing, Vol.76(3), 367-392.

Gary J. Russel, Wagner A. Kamakura. (1997). Modeling Multiple Category Brand Preference with Household Basket Data. Journal of Retailing, Volume 73(4), 439-461.

Gregory Piatetsky-Shapiro, William Frawley. (1991). Knowledge Discovery in Databases. AAAI/ MIT Press.

Harald Hruschka, Martin Lukanowicz, Christian Buchta. (1999). Cross-category sales promotion e¤ects. Journal of Retailing and Consumer Services, Volume 6, 99-105.

Jaihak Chung and Vithala R. Rao. (2003). A General Choice Model for Bundles with Multiple-Category Products: Application to Market Segmentation and Optimal Pricing for Bundles. Journal of Marketing Research, Vol. 40, No. 2 , 115- 130.

Jong Soo Park, Ming-Syan Chen, Philip S. Yu. (n.d.). Using a Hash-Based Method with Transaction Trimming and Database Scan Reduction for Mining Association Rules. IEEE Transactions on Knowledge and Data Engineering.

Julander, C.-R. (1992). Basket Analysis: A New Way of Analysing Scanner Data. International Journal of Retail and Distribution Management, Volume 20 (7), 10-18.

KatrinDippold, Harald Hruschka. (2010). Variable Selection for Market Basket Analysis. University of Regensburg Working Papers in Business, Economics and Management Information Systems.

Kumar, N. (2009). Kill a brand, keep a customer. Harvard Business Review.

M. Houtsma and A. Swami. (1993). Set Oriented Mining of Association Rules. San Jose, California: IBM Almaden Research Center.

M.J.Zaki, M.Ogihara, S. Parthasarathy. (1996). Parallel Data Mining for Association Rules on SharedMemory Multi-processors. New York: University of Rochester.

S. Zongyao, Mining local association patterns from spatial dataset, International Conference on Fuzzy Systems and Knowledge Discovery, 2010, pp. 1455 – 1460.

W. J. Pei, Mining Association Rules Based on Apriori Algorithm and Application", International Forum on Computer Science-Technology and Applications, 2009, pp. 141-143.

B. Vo, Mining Traditional Association Rules using Frequent Item Sets Lattice, International Conference on Computers & Industrial Engineering, 2009.

N. Qiang, Association Classification Based on Compactness of Rules, International Workshop on Knowledge Discovery and Data Mining, 2009, pp. 245-247.

Y. J. Wang, A Novel Rule Weighting Approach in Classification Association Rule Mining, IEEE International Conference on Data Mining Workshops, 2007, pp. 271 – 276.

V. Bartik, Association based Classification for Relational Data and its Use in Web Mining, IEEE Symposium on Computational Intelligence and Data Mining, 2009, pp. 252 – 258.

R. Sumithra, Using Distributed Apriori Association Rule and Classical Apriori Mining Algorithms for Grid Based Knowledge Discovery, International Conference on Computing Communication and Networking Technologies, 2010, pp. 1 –5.

❑ ❑ ❑