Product Recommendation

K.S.Anu Kiruthekaa

MCA, Sri Krishna College of Engineering and Technology, Coimbatore.

Abstract—Product recommendation is a part of the world-wide-web and emergence of e-commerce as a platform to define a personalized information retrieval technique used to identify set(s) of items that will be the interest of a certain user. Apriori is an algorithm for learning association rules, the purpose of the Apriori Algorithm is to find associations between different sets of data. It is sometimes referred to as "Market Basket Analysis". Each set of data has a number of items, called a transaction. The output of Apriori is sets of rules that tell us how often items are contained in sets of data. The key idea is that any item set that occurs frequently together must have each item (or any subset) occur at least as frequently.

Keywords— itemsets; apriori; transaction; association; frequent itemset; recommendation.

I. INTRODUCTION

Association rules can be mined and this process of mining the association rules is one of the most important and powerful aspect of data mining. One of the main criteria of ARM is to find the relationship among various items in a database. An association rule is of the form A→B where A is the antecedent and B is the Consequent. And here A and B are item sets and the underlying rule says us purchased by the customers who purchase A are likely to purchase B with a probability percentage factor as %C where C is known as confidence such a rule is as follows:

“Seventy per cent of people who purchase soap will also like to purchase Shampoo” This helps the shop managers to study the behaviour or buying habits of the customers to increase the sales Based on this study items that areregularly purchased by the customers are put under closed proximity.

For example persons who purchase milk will

Also likely to purchase Bread. The interestingness measures like support and confidence also plays a vital role in the association analysis. The support is defined as percentage of transactions that contained in the rule and is given by

Support = (# of transactions involving A and B) / (total number of transactions).

The other factor is confidence it is

Confidence = Probability (B if A) = P(B/A)

Confidence = (No. of transactions involving A and B) / (total number of transactions that have A).

Consider the following example as

CUSTOMER	ITEM PURCHASED	ITEM PURCHASED
1	MILK	BREAD
2	TEA	MILK
3	SOAP	SHAMPOO
4	COFFEE	CAKE

If A is “purchased Milk“ and B is “purchased Bread” then Support=P(A and B)=1/4

Confidence=P(B/A)=1/2

Item sets that satisfy minimum support and minimum

Confidence are called strong association rules.

II. ASSOCIATION RULE

Let I = {i1,i2,…,in} be set of items (n binary attributes).Let D = {t1,t2,…,tn} be a database having a set oftransactions where each transaction in D has an uniquetransaction T and contains a subset of the items in I. The support of rule X and Y are called antecedent (LHS:left hand side) and consequent (RHS: right hand side) ofthe rule. Consider in Figure 1, the set of items is I ={Breads, Butter, Juices, Dairy, Canned Foods}. Anexample rule for the supermarket could be {Breads,Butter}↔{Juices} meaning that if Breads and Butter isbought, customers also buy Juices. The strength of anassociation rule can be measured in terms of its supportand confidence. Support determines how often a rule isapplicable to a given data set, while confidencedetermines how frequently items in Y appear intransactions that contain X. The formal definitions ofsupport, confidence, and lift are defined

$\text{Support}(X \rightarrow Y) =$	$\dfrac{\text{Probability}(X \cup Y)}{\text{Total number of transactions}}$	(1)
$\text{Confidence}(X \rightarrow Y) =$	$\dfrac{\text{Probability}(X \cap Y)}{\text{Number of transaction}(X)}$	(2)

$\text{Lift}(X \rightarrow Y) =$	$\dfrac{\text{Probability}(X \cap Y)}{\text{Probability}(X)\text{Probability}(Y)}$	(3)

III. APRIORI ALGORITHM

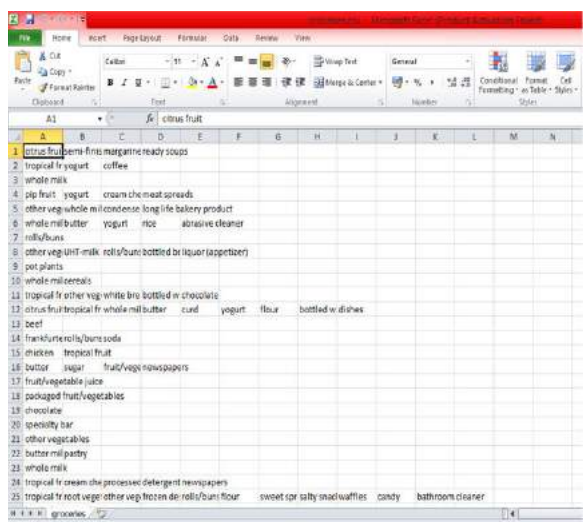
Apriori algorithm is one of the Data Mining algorithmwhich is used to find the frequent items/ itemsets from agiven data repository. The algorithm mainly involves 2steps: Pruning and joining . The Apriori property is theimportant factor to be considered before proceeding withthe algorithm.Apriori Property: if an item X is joined with item Y, Support(X U Y) = min(Support(X), Support(Y))

IV. R PROGRAMMING

In order prove the above association rule by using apriori algorithm R programming is being used as the coding language.

- Software Needed: R and R studio
- Packages:

Arules and ArulesViz Packages for implementing apriori algorithm Dataset used:Groceries dataset which consists of many item transaction of a single shop



The above dataset consists of 9835 transactions (rows) and 169 items (columns).

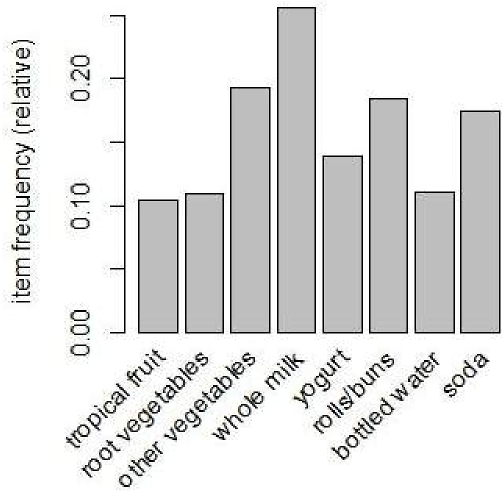
Importing dataset and inspecting items

```
> library(arules)
> library(datasets)
> data("Groceries")
> Groceries
transactions in sparse format with
 9835 transactions (rows) and
 169 items (columns)
> inspect(Groceries[1:3])
    items
[1] {citrus fruit,semi-finished bread,margarine,ready soups}
[2] {tropical fruit,yogurt,coffee}
[3] {whole milk}
```

The above fig represents the inspection of first three transaction in the imported dataset.

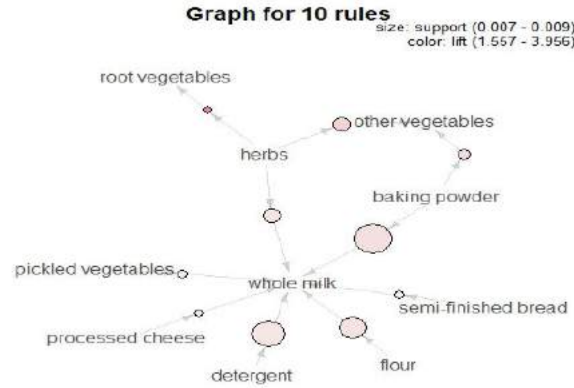
Plotting

itemFrequencyPlot(Groceries,support=0.10



Recommended products

```
Console
> inspect(m1)
     lhs             rhs                       support    confidence lift
[1]  {whole milk} => {other vegetables}        0.07483477 0.29287704 1.5136341
[2]  {whole milk} => {rolls/buns}              0.05663447 0.22164743 1.2050318
[3]  {whole milk} => {yogurt}                  0.05602440 0.21925985 1.5717351
[4]  {}           => {other vegetables}        0.19349263 0.19349263 1.0000000
[5]  {whole milk} => {root vegetables}         0.04890696 0.19140470 1.7560310
[6]  {}           => {rolls/buns}              0.18393493 0.18393493 1.0000000
[7]  {}           => {soda}                    0.17437722 0.17437722 1.0000000
[8]  {whole milk} => {tropical fruit}          0.04229792 0.16553920 1.5775950
[9]  {whole milk} => {soda}                    0.04006101 0.15678472 0.8991124
[10] {}           => {yogurt}                  0.13950178 0.13950178 1.0000000
[11] {whole milk} => {bottled water}           0.03436706 0.13450060 1.2169396
[12] {whole milk} => {pastry}                  0.03324860 0.13012336 1.4625865
[13] {whole milk} => {whipped/sour cream}      0.03223183 0.12614405 1.7597542
[14] {whole milk} => {citrus fruit}            0.03050330 0.11937923 1.4423768
[15] {whole milk} => {pip fruit}               0.03009659 0.11778750 1.5570432
[16] {whole milk} => {domestic eggs}           0.02999492 0.11738957 1.8502027
[17] {whole milk} => {sausage}                 0.02989324 0.11699164 1.2452520
[18] {}           => {bottled water}           0.11052364 0.11052364 1.0000000
[19] {}           => {root vegetables}         0.10899847 0.10899847 1.0000000
[20] {whole milk} => {butter}                  0.02755465 0.10783924 1.9460530
[21] {whole milk} => {newspapers}              0.02735130 0.10704337 1.3411103
[22] {}           => {tropical fruit}          0.10493137 0.10493137 1.0000000
[23] {whole milk} => {fruit/vegetable juice}   0.02663955 0.10425786 1.4421604
[24] {whole milk} => {curd}                    0.02613116 0.10226821 1.9194805
[25] {whole milk} => {brown bread}             0.02521607 0.09868683 1.5212930
[26] {}           => {shopping bags}           0.09852567 0.09852567 1.0000000
[27] {whole milk} => {shopping bags}           0.02450432 0.09590131 0.9733637
[28] {whole milk} => {margarine}               0.02419929 0.09470752 1.6170980
[29] {}           => {sausage}                 0.09395018 0.09395018 1.0000000
[30] {}           => {pastry}                  0.08896797 0.08896797 1.0000000
[31] {whole milk} => {pork}                    0.02216573 0.08674891 1.5047187
[32] {whole milk} => {beef}                    0.02125064 0.08316753 1.5851795
[33] {}           => {citrus fruit}            0.08276563 0.08276563 1.0000000
[34] {}           => {bottled beer}            0.08052872 0.08052872 1.0000000
[35] {whole milk} => {frankfurter}             0.02053889 0.08038201 1.3630295
>
```



V. CONCLUSION

A data set with 9835 transactions (rows) and 169 items (columns) has been used. A set of association rules are obtained by applying Apriori algorithm.By analysing the data,and giving different support and confidence values,we can obtain different number of rules. This helps to improve the customer shopping experience,increase product sales In future the same algorithm can be modified and it can be extended in the future work, which also decreases the time complexity.

References

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

[2] Jiawei Han and Micheline Kamber, “Data Mining: Concepts and Techniques”, 2 edition (2006)

[3] Gordon S. Linoff and Michael J. Berry, “Data Mining Techniques: For Marketing, Sales, and Customer Relationship Management”, 3rd Edition edition (1 April 2011)

[4] Agrawal, R. and J. Shafer, “Parallel Mining of Association Rules”, IEEE Transactions On Knowledge And Data Engineering, Vol. 8, 1996.

[5] Trnka., “Market Basket Analysis with Data Mining Methods”,International Conference on Networking and Information

Technology (ICNIT) ,2010

[6] Hands-On Programming with R,By Garrett Grolemund Publisher: O'Reilly Media (June 2014)