Identify the changing ways and means of marketdata using Mining Association Rule

EKBAL RASHID
Postdoc Researcher,
Technical University of Sofia, Sofia, BULGARIA

NIKOS E. MASTORAKIS
Sector of Electrical Engineering and Computer Science, Hellenic Naval Academy, Piraeus, GREECE
and English Language Faculty of Engineering, Technical University of Sofia, Sofia, BULGARIA

Abstract: -In this research paper author is using the mining association rule learning or affinity analysis which is also called Market Basket Analysis (MBA). As people knew if somebody buys certain items then it is more likely to buy another set of items. This technique makes use of different data mining algorithms and provides a proper analysis on the data about the items purchased. The analysis is therefore recorded as a report and the report is provided to the end user. The output of market basket analysis is generally a set of rules that we can use to make business decision. Simply market basket analysis looks at the purchase coincidence with the items purchase among transaction and what product is purchased with what product. Therefore it helps the end user to determine the sales of an item based on the report submitted. Example: If you buy milk then you are more likely to buy bread or butter.

Keywords: *Data Mining; Market Basket Analysis; Association Rules; Business*

Received: April 9, 2024. Revised: July 15, 2024. Accepted: August 18, 2024. Published: September 23, 2024.

1. Introduction

Market basket analysis is currently a best analysis method with which the end users can be benefited in making business strategic decisions [1][2][3][4]. The existing system mainly makes use of association rules and apriori algorithm to analyze small amounts of data [5][6]. Many researchers and scientists have provided different methods and also introduced additional algorithms to apriori algorithm in order to achieve the desired results in an appropriate time [7][8][9][10][11][12]. However there exists certain draw backs which may affect the end result. Some drawbacks are listed below:

- Complexity grows exponentially.
- Scans the database many times and generate many association rules.
- Time consumed is more.

2. Proposed System

The proposed system mainly concentrates on providing the efficient output. So, in order to achieve this we implement algorithms like Apriori and Association [13]. The Apriori algorithm is useful to generate the frequent item set by using minimum support count i.e., MIN_SUP. Association rule mining is used to generate the strong rules by using rule(s□l- s) and minimum confidence i.e., MIN_CONF [14][15] .

2.1 Steps

Following steps are required for data analysis.

- Taking the dataset
- If the dataset is not in the correct format then data preparation and cleaning is done in order to remove andmodify corrupted, inaccurate data.
- Exploratory data analysis (EDA) is done in order to identify the dataset easily in terms of histograms, plotsetc.
- Feature extraction is creation of new set of features from original raw data.
- Apriori and Association algorithms are applied by setting support and confidence values.
- Strong rules are generates.
- Storing data in database.

2.2 Advantages

There are many advantages while applying the proposed system which has given below.

- Increasing profits in business strategies.
- Complexity reduces.
- Better decision is taken.
- It consumes less time.

3. Proposed Algorithm

In this section we are implementing by using Apriori and Association algorithm in a tabular form. See table

Table 1

Tid	ITEMSETS
1	Bread, Milk
2	Bread, Rice, Eggs, Jam
3	Milk, Rice, Jam, Coke
4	Bread, Milk, Rice, Jam
5	Bread, Milk, Rice, Coke

3.1 ITEMSET:

It is defined as the collection of zero or many items in Association analysis

Itemset={Bread,Milk}

3.2 SUPPORT COUNT:

Number of transactions that contains all the items in Association rule

Ϭ = Σ ti

If we take {Bread, Milk} then we have to analyze if this pair exists or notϬ = Σ ti .. (Bread, Milk) =3

3.3 SUPPORT:

It defines "how often the rule occurred" is defined as the measure of rule support.

It is the difference between the support count and the number of transaction or purchases

S = Ϭ (XU Y)/N

S = 3/5

=0.6

3.4 CONFIDENCE:

It determines how frequently items in „Y" appears in the transaction that contain „X".

CONFIDENCE = Ϭ(XUY)/Ϭ(X)

p(Y/X)

X Y

If „X" then there is „Y"

Bread MilkX Y

= Ϭ (Bread U Milk) / Ϭ (Bread)

= 3/4

= 0.75

Confidence(X Y)

P(Y/X) = supcnt (XU Y) /supcnt (X)

4. Frequent itemset generation in apriori

Taking, MIN_SUP=60% and 5 transactions.

Table 2

Tid	ITEMSET
1	M O N K E Y
2	D O N K E Y
3	M A K E
4	M U C K Y
5	C O O K I E

I = {A, C, D, E, I, K, M, N, O, U, Y}

*MIN_SUP = 60%

= 60/100 *5 = 3

If the support count value of item is less than the given support count value that is "3" then we to prune thatitem sets.

4.1 Candidate 1 itemset

Table 3

ITEM	Ϭ
A	1
C	2
D	1
E	4
I	1
K	5
M	3
N	2
O	3
U	1
Y	3

RED indicates pruned candidate1-itemsets.

4.2 Candidate 2 itemset

Table 4

ITEM	б
EK	4
EM	2
EO	3
EY	2
KM	3
KO	3
KY	3
MO	1
MY	2
OY	2

RED indicates Pruned candidate 2-itemsets.

s→l-s

4.3 Candidate 3 itemset

Table 5

ITEM	б
EKO	3
EKM	2
EKY	2
KOM	1
KOY	2
KMY	2

RED indicates pruned candidate 3-itemsets.

"EKO" is left in candidate 3-itemset after pruning all itemsets which are less than support count value "3".So, the final frequent itemset generated in APRIORI algorithm is {EKO}.

5. Association Rule Generations

There are two processes to construct association rule generation.

MIN_SUPPORT MIN_CONFIDENCE
*Given MIN_CONF=80%;

In previous, frequent item set generated is {EKO} that is l = {E, K, O} Consider all non-empty subsets ---------- {E}, {K}, {O}, {E, K}, {E, O}, {K, O}
If all the subsets are frequent then obtained frequent itemset {E, K, O} also be frequent by using the rule.

{E} -- {K, O}

{K} -- {E, O}

{O} -- {E, K}

{K, O} -- {E}

{E, O} -- {K}

{E, K} -- {O}

By using the formula,

{X} □ {Y} = supcnt (XU Y)/supcnt (X).

*{E} -- {K, O} = supcnt{E, K, O}/supcnt{E} = 3/4 = 75%.

*{K} -- {E, O} = supcnt{K, E, O}/supcnt{K} = 3/5 = 60%.

*{O} -- {E, K} = supcnt{O, K, E}/supcnt{O} = 3/3 = 100%.

*{K, O} -- {E} = supcnt{K, O, E}/supcnt{K, O} = 3/3 = 100%.

*{E, O} -- {K} = supcnt{E, O, K}/supcnt{E, O} = 3/3 = 100%.

*{E, K} -- {O} = supcnt{E, K, O}/supcnt{E, K} = 3/4 = 75%.Strong rules identified are
{O} -- {E, K}, which is 100%

{K, O} -- {E}, which is 100%

{E, O} -- {K}, which is 100%

6. Conclusion

By using two algorithms apriori and association algorithms we can easily generate frequent item sets from large amounts of data, so that we are getting results as easy and little amount of data from large

data sets.This frequent item sets are used for making business decisions easily in small amount of time for getting profits.

References:

[1]. Sheenu Verma, Sakshi Bhatnagar. "An Effective Dynamic Unsupervised Clustering Algorithmic Approach for Market Basket Analysis" .International Journal of Enterprise Computing and Business Systems 2014:4(2). [2]. AA Raorane, RV Kulkarni, BD Jitkar. "Association Rule – Extracting Knowledge Using Market Basket Analysis". Research Journal of Recent Sciences 2012:1(2):19-27.

[3]. Maurizio Marek. "Data Mining Concepts and Techniques.", E-Commerce Winter 2011

.[4]. A Herman, LE Forcum, Harry Joo "Using Market Basket Analysis in Management"

.Journal of Management 2013:39(7):1799-1824.

[5]. Saurkar Anand V, Bhujade V, Bhagat P, Khaparde A. "A Review Paper on various Data Mining Techniques" .International Journal of Advanced Research in Computer Science and Software Engineering 2014:4(4):98-101.

[6]. Paramjit Kaur, Attwal Kanwalpreet S. "Data Mining: Review" .International Journal of Computer Science and Information Technologies 2014:5(5):6225-6228.

[7]. Wu X, Kumar V, Quilan JR., Ghosh J, Yang Q, Motoda H. "Top 10 Algorithms in Data Mining"

.Springer-Verlay London Limited 2007:14:1-37.

[8]. EWT Ngai, Xiu Li, Chau DCK. "Application of Data Mining Techniques in Customer Relationship Management: A Literature Review and Classification" .Elsevier-Expert Systems with Applications 2009:36:2592-2602.

[9] . M Ramageri Bharati. "Data Mining Techniques and Applications.", Indian Journal of Computer Science and Engineering:1(4):301-305.

[10]. Zhao Quiankun, Bhowmick Sourav. "Association Rule Mining: A Survey" .Technical Report CAIS Nanyan Technological University, Singapore 2003:1-20.

[11]. Trnka Andrej. "Market Basket Analysis with Data Mining Methods" .International Conference on Networking and Information Technology 2010:446-450.

[12]. A Tejaswini Hilage, RV Kulkarni. "Application of data mining techniques to a selected business organization with special reference to buying behavior" .International Journal of Database Management Systems 2011:3(4):169-181.

[13]. Savi Gupta, Roopal Mamtora. "A Survey on Association Rule Mining in Market Basket Analysis"

.International Journal of Information and Computational Technology 2014:4(4):409-414.

[14]. H Kargupta, Han J, Yu Philips S, Motwani R, Kuman Vipin. Book: Next Generation of Data Mining.CRC Press, Taylor and Francis Group, LLC 2008.

[15]. Dipti Verma, Nashine Rakesh. Data Mining: Next Generation Challenges and Future Directions.International Journal of Modeling and Optimization 2012:2(5):603-608.

Research