Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 7, 2021, Pages. 51- 59
Received 05 May 2021; Accepted 01 June 2021.

Market Basket Analysis: Identify the changing trends of market datausing association rule mining

DEVENDRA KUMAR[1], RAUSHAN KASHYAP[2], PROF. N. GAYATHRI[3]

[1]School of Computer Science and Engineering Galgotias University Greater Noida, India
[2]School of Computer Science and Engineering Galgotias University Greater Noida, India
[3]School of Computer Science and Engineering Galgotias University Greater Noida, India

Abstract

Market Basket Analysis (MBA) also known as association rule learning or affinity analysis, is a data mining technique that can be used in various fields, such as marketing, bioinformatics, education field, nuclear science etc. The main aim of MBA in marketing is to provide the information to the retailer to understand the purchase behavior of the buyer, which can help the retailer in correct decision making. There are various algorithms are available for performing MBA. The existing algorithms work on static data and they do not capture changes in data with time. But proposed algorithm not only mine static data but also provides a new way to take into account changes happening in data. This paper discusses the data mining technique i.e. association rule mining and provide a new algorithm which may helpful to examine the customer behaviour and assists in increasing the sales.

Keywords: Data Mining; Association rule; Score Table; Outliers

1. Introduction

Today, the large amount of data is being maintained in the databases in various fields like retail markets, banking sector, medical field etc. But it is not necessary that the whole information is useful for the user. That is why, it is very important to extract the useful information from large amount of data. This process of extracting useful data is known as data mining or A Knowledge Discovery and Data (KDD) process. The overall process of finding and interpreting patterns from data involves many steps such as selection, preprocessing, transformation, data mining and interpretation.[1,2,3]

Data mining helps in the business for marketing. The work of using market basket analysis in management research has been performed by Aguinis et al.[1] Market basket analysis is also known as association rule mining. It helps the marketing analyst to understand the behavior of customers e.g. which products are being bought together. There are various techniques and algorithms that are available to perform data mining.[4]

1.1. Techniques of Data Mining

There are many data mining techniques and algorithms are available to discover meaningful pattern and rules.
These techniques have been discussed by Saurkar et al.[5].
There are many different techniques are as follow:

Classification: In classification, first examine the features of newly presented object and assign it to a predefined class for example classify the credit applicants as low, medium or high risk.[5]

Association: The main goal of association is to establish the relationship between items which exist in the market. The typical examples of association modeling are Market basket Analysis and cross selling programs. The tools used for association rule mining are apriori algorithm and weka tool kit.[6,7,8]

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 7, 2021, Pages. 51- 59
Received 05 May 2021; Accepted 01 June 2021.

Prediction: In this functionality, prediction of some unknown or missing attributes values based on other Information. For example: Forecast the sale value for next week based on available data.[8,9]

Clustering: In this, Data Mining organizes data into meaningful sub-groups (clusters) such that points within the group are similar to each other, and as different as possible from the points in the other groups. It is an unsupervised classification. An effective dynamic unsupervised clustering algorithmic approach for market basket analysis has been proposed by Verma et al.[2].

Outlier Analysis: In this, Data Mining is done to identify and explain exceptions. For example, in case of Market Basket Data Analysis, outlier can be some transaction which happens unusually. [10]

1.2. Association Rule Mining

Association rule mining is useful for discovering interesting relationships hidden in large data sets. In the following example, there are some transactions of the shop have been taken as shown in Table 1.

Table1. An example of Market Basket Transactions.

Transaction ID (TID)	Items
	Butter, Cheese, Burger
	Milk, Cheese, Butter
	Butter, Milk

The Interesting relationships can be represented in the form of association rules as shown below:
Milk Butter

The above rule shows that there is a strong relationship between milk and butter. It shows that many customers buy milk and butter together. These rules can be helpful for retailers to understand buying nature of customers.

One of the most popular data mining approaches is to find frequent item sets from a transaction dataset and derive association rules. [7]

The survey on association rule mining has been performed by Zhao et al.[11]. In this survey, different types of mining such as association rule mining, classification, clustering and other techniques have been discussed. Further two basic measures have been discussed for association rules i.e. support and confidence.

In this researchknowledge has provided about Apriori series approaches, AIS algorithm, Apriori Algorithm, FP-Tree Algorithm(Frequent Pattern-Tree Algorithm), RARM(Rapid Association rule Mining) Algorithm. But from all these algorithms, Apriori is the biggest improvement from previous algorithms and easy to implement.

The work of market basket analysis with data mining methods has been proposed by Andrej . Market basket analysis had been implemented based on Six Sigma methodology. The aim of this study was to improve the result and change the sigma performance level of the process. General rule induction (GRI) algorithm was used in this study to establish the association rules.

Hilage et al. has proposed an application of data mining techniques to a selected business organization with special reference to buying behavior. The result was examined after applying association rule mining technique, rule induction technique and apriori algorithm. Subsequently the results of these three techniques were combined and efforts were made to understand the correct buying behavior of the customer.

The work of extracting knowledge using market basket analysis has been proposed by Raorane et

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 7, 2021, Pages. 51- 59
Received 05 May 2021; Accepted 01 June 2021.
al.. Association rule data mining technique was used. For this they used the dataset of supermarket and analyse the daily transactions of the market. The main purpose of this study was to arrange the products of supermarket in such a way so that the profit of supermarket may increase.

The existing work for association rule mining in market basket analysis is MBA in Large Database Network, MBA in Multiple Store environment, MBA using Fast Algorithm.

1.3. Outlier Detection

According to Hawkins who defines "an outlier is an observation which deviates so much from other observations as to arouse suspicions that it was generated by a different mechanism".

The work of FP-Outlier-Frequent pattern based outlier detection has performed by He et al.. A new method of outlier detection by discovering frequent pattern from the data set was proposed. A measure called FPOF (Frequent pattern outlier factor) to detect the outlier transactions has defined and proposed the FindFPOF algorithm to discover outliers.

The work of outlier detection of Business Intelligence using data mining technique has implemented by Khan et al.Before this work, the main focus of researchers was to found pattern from large datasets which may help in decision making. However outlier detection was not the main focused area of research. Hence this work was the advancement in outlier detection.

Although data mining has become popular as an emerging technique, still there are several issues to be resolved to make it useful in diverse domains. Some of the issues faced by data mining are quality of data, inter- operability, security and privacy etc. The major issue with the data mining is its lack of taking into account the analysis of real time data. To follow the changing trends of data, periodic mining come in existence. Periodic mining refers to perform the data mining after fixed time period. For example, a departmental store mines for association rules every quarter to discover current purchase behavior of customers.

2. Existing Algorithm

There are many algorithms are available for association rule mining. Existing algorithms work on the static data. They find the good association rules on basis of various metrics such as support, confidence, lift etc. In these algorithms, when next time they perform data mining, then algorithm automatically does not capture the changes in data. That is why they use some another comparison algorithm to track the change in data.

3. Proposed Algorithm

Our proposed algorithm also performs association rule mining. It works on change modeling concept. Basically, change modeling is used to understand the dynamics of data generation process by examining changes that have taken place in discovered patterns. It works on the dynamic data and performs periodic mining. Periodic Mining is actually the mature usage of KDD process.

3.1. ARM-Predictor Algorithm

This algorithm is trying to capture the changing trends of transactions in Market Basket Analysis. It is based on the basic idea of collaborating Association Rule Miner, Changes in Association Rule Predictor based on some logic to get the strong relationship between the various attributes (i.e the goods placed in market). The main thrust is on finding the association between various items in transactions. We keep track on the items which are associated with high confidence (i.e $X \rightarrow Y$, then confidence = $n(X \cap Y)/n(x)$). So result of this algorithm will be two sets of association rules:

1. Association rules which are highly predictable for future windows.

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 7, 2021, Pages. 51- 59
Received 05 May 2021; Accepted 01 June 2021.
2. Outliers (Association Rules which are least probable to come in next windows).
Input: Set of Transactions Output: Predicted Association Rules, Outdated Association Rules

3.1.1. Definitions and Specifications

Support (X): Support of item is the number of times an item occurs in transactions in a database.
Confidence: Confidence is a term associated with association rule, It is defined mathematically as : Confidence =Support(X∩Y)/Support(X)
Score (X→Y): It is the value which is assigned to attributes of association on the basis of confidence of that
association rule as shown in Table 2.

Table 2:Score assignment based on their confidence

Confidence in %age	Score Assigned
<=10	0
>10 and <=20	1
>20 and <=30	2
>30 and <=40	4
>40 and <=60	6
>60 and <=90	8
>90	9

3.1.2. Data Set

For this algorithm to run, the data set had been taken from Extended bakery datasets and store it in 4 windows and algorithm work on 2000 transactions in each window and 26 items, items can be extended up to n. (Link to website
: https://wiki.csc.calpoly.edu/ datasets/wiki/apriori)

3.1.3. Stages of Algorithm

Stage 1 : In the first stage, we are having with us binary datasets of 4 windows with specifications as explained in
previous section.
Apriori Algorithm : In this part we just run the apriori algorithm on the binary datasets of all the windows and foundfrequent itemsets and further association rules from them.
Stage 2 : This stage can be divided in two sub-parts in which two algorithms are run alternatively.
Part 1 - ARM-Update : This algorithm creates Score Table and the structure is shown in Fig.1(a) and then updatingscore table as the data from consecutive windows come.

```
ARM-Update( Windowi, ConfidenceToScoreTable, ScoreTable )
{
For ( i = start-of-Windowi ; i < end-of-Windowi ; i ++)
{
```

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 7, 2021, Pages. 51- 59
Received 05 May 2021; Accepted 01 June 2021.

N = AssignScore (ith assoction rule, ConfidenceToScoreTable) ; CreateEntryScoreTable(N,ith association rule);
}
}
where

AssignScore (ith association rule, ConfidenceToScoreTable) : It is a function which is taking Input some Association rule and Confidence to ScoreTable and this algorithm is used with algorithm with Part-2 Algorithm and supplies Information processed to Part-2 algorithm which further processes the information.
CreateEntryScoreTable (N,ith association rule) : It is a function which create new entry in the score table if some ith association rule is not in the ScoreTable or if present, then just add score N to existing rule.
Part 2 - ARM-Predictor : This part is run after we have run ARM-Update Algorithm, this algorithm find the outliers on the basis of some threshold value.
ARM-Outlier (ScoreTable) {
for (i=0;i$<$number of months ;i++) {
A = FindUpperRules(Rules above threshold);
B = FindLowerRules(Rules below threshold) //containing outliers ;
}
FindUpperRules() : It is a function which is finding set of association rules above threshold value as shown below in
Table 4.
FindLowerRules() : This algorithm find the set of association rules below threshold value as shown below in Table
5. These rules are called as outlier.

3.1.4. Experimental results

3.1.5. (a) Specifications

ARM-Update Algorithm
Input : Window, Look-Up Table Output : Score Tablewhere
Window: It contains association rules for some particular time period Look-Up Table: It contains Confidence to corresponding Score Values
Score Table : Association rules along rows and their attributes in columns with their scoresARM-Predictor Algorithm
Input : Score Table Output : OutliersWhere
Score Table : Association rules along rows and their attributes in columns with their scores
Outliers: Set of Association Rules which are above score-threshold, Set of Association Rules which are below score-threshold

3.1.5. (b) Points for analyzing results

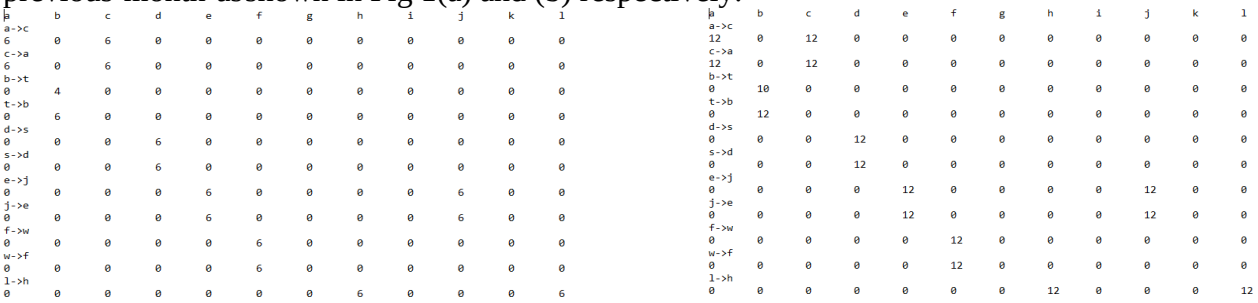
The results below are shown in order as follows:
1. In the Score Table as shown in Fig 1(a), the attributes are set aside row wise at the top and named them a,b,c andso on for the simplicity of the transaction. a,b,c and so on are the items which are kept in market basket. As shown in the Fig. 1(a), after the top row of total number of items, there are association rules with their assigned score.

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 7, 2021, Pages. 51- 59
Received 05 May 2021; Accepted 01 June 2021.

2. Upper Association Rules, meaning rules which are above threshold are being printed.
3. Lower Association Rules, meaning the association rules which are below threshold.

3.1.5. (c) Results with data set

i) Score Table after first month and after second month which follows the change in data from previous month asshown in Fig 1(a) and (b) respectively.

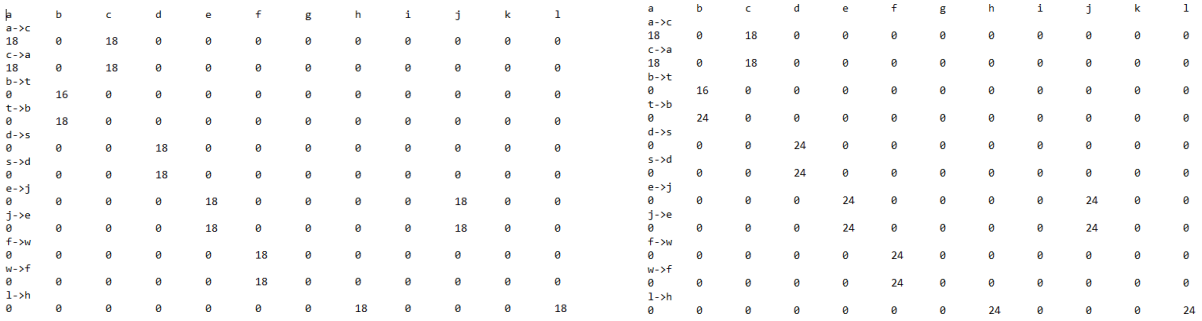


a	b	c	d	e	f	g	h	i	j	k	l
a->c											
6	0	6	0	0	0	0	0	0	0	0	0
c->a											
6	0	6	0	0	0	0	0	0	0	0	0
b->t											
0	4	0	0	0	0	0	0	0	0	0	0
t->b											
0	6	0	0	0	0	0	0	0	0	0	0
d->s											
0	0	0	6	0	0	0	0	0	0	0	0
s->d											
0	0	0	6	0	0	0	0	0	0	0	0
e->j											
0	0	0	0	6	0	0	0	0	6	0	0
j->e											
0	0	0	0	6	0	0	0	0	6	0	0
f->w											
0	0	0	0	0	6	0	0	0	0	0	0
w->f											
0	0	0	0	0	6	0	0	0	0	0	0
l->h											
0	0	0	0	0	0	0	6	0	0	0	6

a	b	c	d	e	f	g	h	i	j	k	l
a->c											
12	0	12	0	0	0	0	0	0	0	0	0
c->a											
12	0	12	0	0	0	0	0	0	0	0	0
b->t											
0	10	0	0	0	0	0	0	0	0	0	0
t->b											
0	12	0	0	0	0	0	0	0	0	0	0
d->s											
0	0	0	12	0	0	0	0	0	0	0	0
s->d											
0	0	0	12	0	0	0	0	0	0	0	0
e->j											
0	0	0	0	12	0	0	0	0	12	0	0
j->e											
0	0	0	0	12	0	0	0	0	12	0	0
f->w											
0	0	0	0	0	12	0	0	0	0	0	0
w->f											
0	0	0	0	0	12	0	0	0	0	0	0
l->h											
0	0	0	0	0	0	0	12	0	0	0	12

Fig. 1 Score table (a) after first month and (b) second month transaction

iii) Score table after third month which follows the change in data from second month and after fourth month which follows the change in data from third month as shown in Fig 2 (a) and (b) respectively.



a	b	c	d	e	f	g	h	i	j	k	l
a->c											
18	0	18	0	0	0	0	0	0	0	0	0
c->a											
18	0	18	0	0	0	0	0	0	0	0	0
b->t											
0	16	0	0	0	0	0	0	0	0	0	0
t->b											
0	18	0	0	0	0	0	0	0	0	0	0
d->s											
0	0	0	18	0	0	0	0	0	0	0	0
s->d											
0	0	0	18	0	0	0	0	0	0	0	0
e->j											
0	0	0	0	18	0	0	0	0	18	0	0
j->e											
0	0	0	0	18	0	0	0	0	18	0	0
f->w											
0	0	0	0	0	18	0	0	0	0	0	0
w->f											
0	0	0	0	0	18	0	0	0	0	0	0
l->h											
0	0	0	0	0	0	0	18	0	0	0	18

a	b	c	d	e	f	g	h	i	j	k	l
a->c											
18	0	18	0	0	0	0	0	0	0	0	0
c->a											
18	0	18	0	0	0	0	0	0	0	0	0
b->t											
0	16	0	0	0	0	0	0	0	0	0	0
t->b											
0	24	0	0	0	0	0	0	0	0	0	0
d->s											
0	0	0	24	0	0	0	0	0	0	0	0
s->d											
0	0	0	24	0	0	0	0	0	0	0	0
e->j											
0	0	0	0	24	0	0	0	0	24	0	0
j->e											
0	0	0	0	24	0	0	0	0	24	0	0
f->w											
0	0	0	0	0	24	0	0	0	0	0	0
w->f											
0	0	0	0	0	24	0	0	0	0	0	0
l->h											
0	0	0	0	0	0	0	24	0	0	0	24

Fig. 2. Score Table (a) after third month and (b) after fourth month

Outlier Detection

iv) After the fourth month rules as shown in Table 3 we perform outlier detection, at the threshold value = 20

then it divide the rules into two parts upper association rule as shown in Table 4 and lower association rules asshown in Table 5. Lower association rules are known as outliers.

Table 3.Fourth month association rules

Association Rules	Score Assigned	
	A	C
a->c	18	18
	C	C
c->a	18	18
	B	T
b->t	16	16

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 7, 2021, Pages. 51- 59
Received 05 May 2021; Accepted 01 June 2021.

	T	B
t->b	24	24

	D	S
d->s	24	24
	S	D
s->d	24	24
	E	J
e->j	24	24
	J	E
j->e	24	24
	F	W
f->w	24	24
	W	F
w->f	24	24
	L	H
l->h	24	24
	P	H

Table 4. Upper Association rules

Association Rules	Score Assigned	
	T	B
t->b	24	24
	D	S
d->s	24	24
	S	D
s->d	24	24
	E	J
e->j	24	24
	J	E
j->e	24	24
	F	W
f->w	24	24
	W	F
w->f	24	24
	L	H
l->h	24	24
	P	H
p->h	24	24

Table 5. Lower association rules (Outliers)

Association Rules	Score Assigned	
	A	C
a->c	18	18

	C	C
c->a	18	18
	B	T
b->t	16	16
	X	Y
x->y	0	0

4. Conclusion

At present many data mining algorithms have been developed and applied on variety of practical problems. However periodic mining is a new approach in data mining which has gained its significance these days. This field is evolving due to needs in different applications and limitations of data mining. This would enhance the power of existing data mining techniques. Finding out the patterns due to changes in data is in itself an interesting area to be explored. It may helpful in

- Find out interesting patterns from large amount of data.
- Automatically track the changes in facts from previous data; due to this feature it may be helpful in frauddetection.
- Predicting future association rules as well as gives us right methodology to find out outliers.

Authors suggested that, some areas are still there which need to be focused on. Firstly, results have influenced greatly by the manual threshold values for score, so it is needed to automate the threshold values for better recognition of outliers. Secondly, this approach is specifically targeted at Market Basket Data, it may perhaps be extended to other areas.

5. References

1. Raorane AA, Kulkarni RV, Jitkar BD. Association Rule – Extracting Knowledge Using Market Basket Analysis.Research Journal of Recent Sciences 2012:**1(2)**:19-27.

2. Verma Sheenu, Bhatnagar Sakshi. An Effective Dynamic Unsupervised Clustering Algorithmic Approach for Market Basket Analysis.International Journal of Enterprise Computing and Business Systems 2014:**4(2)**.

3. Maurizio Marek. Data Mining Concepts and Techniques. E-Commerce Winter 2011.

4. Herman A, Forcum LE, Joo Harry. Using Market Basket Analysis in Management Research.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. Kaur Paramjit, 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. Ngai EWT, 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. Ramageri Bharati M. Data Mining Techniques and Applications. Indian Journal of Computer Science and Engineering:**1(4)**:301-305.

10. Hawkins D. Identification of Outliers.Chapman and Hall 1980.

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

12. Trnka Andrej. Market Basket Analysis with Data Mining Methods.International Conference on Networking and Information Technology

2010:446-450.

13. Hilage Tejaswini A, Kulkarni RV. 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.

14. Gupta Savi, Mamtora Roopal. A Survey on Association Rule Mining in Market Basket Analysis.International Journal of Information and Computationa Technology 2014:**4(4)**:409-414.