

SURVEY ON MARKET BASKET ANALYSIS USING DATA MINING AND DEEP LEARNING TECHNIQUES

S. Naganandhini, R. Revathi*

ABSTRACT

Market Basket Analysis (MBA) helps to scrutinize the purchase / sale frequency and pattern, trend of the business, market potential and also ascertain the association between products in the customer's purchase. Different kind of rule mining algorithms is used to determine the purchase frequency, purchase pattern for the products and forecasting the sale of products in the business. MBA is a one kind of data mining technique and allows to generate the rules to extract the hidden knowledge from the large data sets like peoples purchase frequency data set. Most of the works are utilized Apriori and FP-Growth algorithms. In recent times, some researchers' utilised neural network based deep learning algorithms to perform MBA. Data Mining based MBA helps to identify the purchasing behaviour and pattern of the customer and assist to increase the sale and provide the better customer service. MBA also used in strategic management, entrepreneurial management and organization behaviour. This survey paper deals with various rule mining algorithms and deep learning algorithms used in MBA by the researchers and summarize the existing works performed in MBA.

Keywords: MBA, Association, Prediction, frequent item set

1. INTRODUCTION

Market basket analysis using data mining and deep learning techniques finds what are the product combinations are purchased together and examine the association between the customer's and products buying behaviour with stock, offers and discounts, displays, advertisements,

branding, demand and supply. Retail outlets / shopping malls / departmental stores/ entertainment places have huge number of customer base due to the more products under one roof and distribute or sell the product in every day. MBA based systems required to analyse and conclude the customer's purchase patterns and frequently purchased products together. It will help to increase the profitability of the business with more customer satisfaction.

MBA is classified into three types based on the data utilized such as predictive, descriptive and differential MBA. Classification and Regression models are utilized in predictive model to determine cross selling like buying additional warranty after purchasing high range smart phones.

Statistics based techniques are employed in descriptive MBA to find out the association between products. Differential MBA employs purchase frequency of the customer between the specific duration and analyse the customer behaviour.

Discovery of association between the products based on association rules plays vital role in MBA based on consumers purchase history. For an example, an association rule is $R1 \Rightarrow R2$, where $R1$ is the predecessor and $R2$ as the subsequent. $R1$ and $R2$ are the product set and it represents, product $R2$ may purchase the consumer who are purchased $R1$ in the probability of confidence percentage. This rule allows to conclude "75% of the peoples who purchase the Masks also purchase Sanitizer". It helps to getting response the queries like "What is sold with Face wash ? or What is sold with Basmathi Rice? or verify the dependency of two products.

¹Department of Computer Science

²Department of Computer Applications

Karpagam Academy of Higher Education, Coimbatore, Tamil Nadu, India

*Corresponding Author



Figure.1. Market Basket Analysis

The figure (1) describes the general process of MBA and example for activities based on the rule mining. Identifying consumer’s purchase frequency and pattern is more vital part of the store management. It helps to understand about consumer behaviour and factors considered while purchase time to the management to perform the sales promotional activity, stock maintenance and display arrangements. MBA helps to identify the most / less frequent items purchased together, fast / slow moving items, consumer’s attitude and behaviour while shopping, and how to arrange the items based on category and purchase pattern. It helps to enhance the sale of multiple products on same time and make it easy to select the products by the customers. In MBA, Association Rule (AR) mining helps to generate rules from the sales details data set based on good support and confidence values. It determines the best rules for MBA. This survey paper deals with various data mining and machine learning algorithms and techniques used by the researchers for MBA.

II. LITERATURE SURVEY

MCChenetal.,[1]integrates the demography and customer behaviour variables with products and food_marttransaction database to find the changes incustomer behaviour.ARwith

Apriori algorithm utilized with 20% confidence and minimum support. 111 rules are generated first period and 78 rules are generated in second period.

Ertek and Demririz[2] utilized AR with Apriori for generating visualization framework forsuper_market data

set and finds the decision making in retailing and relationship between the products and suggestion. This work generated 27rules with2%supportand20% confidence.

MaryamNafari, JamalShahrabi[3] utilized AR with Apriori TdM1 deep learning model in the 4 months retail (supermarket) data set and find the frequentitem sets with0.1supportandconfidence.It helps to product arrangement in shelf with price consideration.

Mutlu YA [4] worked on cloth and accessory data set using apriori algorithm on both in-category and cross categories of product. The association rules and cross categories of product. The AR (8best rules using deep learning) visualized using webgraphics to see relation ship among items.

Zulfikar et al [5] applied MBA to find frequent products using Apriori algorithm at eight clusters of XMART retail company at Indonesia. Generated 10 rules with 12% support and 60%confidence for over 150000 records from 8 clusters of stores.

AbusidaandGultepe[6]designedbusinessstrageyonorde randpurchasedatasetofelecricity company applying apriori algorithm. The association rules generated at high confidencevalueof100%showing frequent demand spare parts.

YNI Sari [7] analysed sales data set ofrestaurent to mine customer intents at minimum supportand 50% confidence. Aenergetic application developed to observe customer purchase patterns within specific time.

Sutisnawatiet al [8] mined relationship between the dining place items and finds AR for frequently and rarely purchased items at minimum confidence 65%.

Ariestya et al[9] analysed small grocery store data set applying AR algorithms FPGrowth and Apriori. The experiment shows that FP growth produced more rules than Apriori at minimum support 45% and confidence 60%

Anggraeni et al[10] analyzed sales data set of electronic store to find most frequently purchased data set of electronic store to find most frequently purchased item by customer. The association rules extracted at minimum item by customer. The association rules extracted at 9% minimum support and 40% confidence 40%.

Motobobo [11], Has[12] and courage[13] proposed hybrid model with AR and ANN for rule generation to mine unique customer behaviour.

Motobobo [11] proposed the model for real time data set. 10 rules are generated using AR with 83% accuracy and 70% of confidence and min. support. 10 rules are generated using ANN with accuracy of 67%.

Has[12] also proposed the same model and generated 9 best rules with above 50% confidence and most buying customer profiles are classified using ANN with 98.733% accuracy.

Courage[13] also proposed AR-ANN model for centralized and also distributed enterprise with 83% of accuracy in real and 100% accuracy in public data set. The data passed to AR to generate output values of confidence and support. ANN utilized these values as input with weights and its sum.

Wanget al.,[14] suggested Classification association rule mining (CARM) with rule weighting technique. It constructs an association rule mining based classifier. This rule weighting and ordering technique helps to select the particular rules and supports confidence. The simulation

results indicates that the rule weighting and ordering techniques yields high accuracy than classification rule mining approach.

Author	Technique Used for MBA	Result	Improvement
MC Chen[1]	AR with Apriori	Developed online system to mine customer patterns in different time. 11 and 78 rules are generated with min support and 20% of confidence	Result visualization and large real data set utilized
Ertok & Demriz [2]	AR with Apriori	Simulated framework for real super_market data set with nearly 88000 transactions for nearly 1600 items. Generated 27 rules with 2% support & 20% confidence	Integrates the association mining and visualization model with 3 D exploration for AR
Maryam N. Jamal S[3]	AR with Apriori & DM1	Utilized 4 months retail (supermarket) data set and Finds the frequent item sets with 0.1 support and confident	Deep Learning model helps to product arrangement in shelf
Mutlu YA[4]	Apriori	From 1 year cloth sales data set, 8 best rules are extracted with 0.05 support and 50% confidence	Deep learning model utilized to define AR rules clearly
WB Zulfikar[5]	Apriori	Generated 10 rules with 12% support and 60% confidence for over 150000 records from 8 clusters of stores	CRISP_DM standard followed on real XMART data set
Abusida[6]	Apriori	Generated 10 best rules with 100% confidence for electricity company data set	Results are simulated with scatter plot for support and confidence values and also visualized 10 AR
YNI Sari[7]	Apriori	Restaurant sales data set utilized and finds the association between season and food	AR simulation with clear results for effects of seasonal x confidence yson foods
Sutisnawati et al [8]	Apriori and FPGrowth	Smallest and Largest frequency item based data set utilised to generate AR with 65% to generate AR with 65%	Season and holidays effect on sale of foods are analysed
Ariestya et al[9]	FP Growth and Apriori	FP Growth with 45% support and 60% confidence for 5 rules and Apriori with same support and confidence for 3 rules	Utilized larger data set with proper dimension
Anggraeni et al[10]	FP Growth and Apriori	10 rules are generated with 9% support and 40% confidence in 1 year electronic sales data	Performance of two algo. compared in large data set
Matobobo[11]	AR with ANN	AR identifies patterns in cent percent accuracy and ANN identifies 75% accuracy	Applied in small data set with simulated results
Has[12]	AR with ANN	9 best rules are generated with 98.73% accuracy	Integrated AR with DL for seasonal and time stamp effects
Courage[13]	AR with ANN	In real data set AR finds 9 patterns and ANN identifies 5 patterns with 83% accuracy	Time is used to find the patterns in AR & ANN
Wang et al.,[14]	Classification association rule mining (CARM) with rule weighting technique	Accuracy of classification with good support and confidence values.	Overlapping problem

III. METHODOLOGY

This paper identified various articles related to MBA published in between 2000 to 2020. It aims to give a review on MBA using AR and DL methods. Most of the works handles identifying product patterns and product recommendation system. Most of the works handles low dimensional data only. When utilizing high dimensional

large purchase or sales transaction data sets, dimensional reduction is needed.

CONCLUSION

Application of DM and ML techniques in MBA is an emerging trend in retail. The existing research works reveals that the Market basket analysis and finding the associations using rule generation plays vital role in commerce. However there are various problems occur in the existing association rules mining systems. Business ontology is one of the vital factors for Market basket analysis. Association rule generation for mining is based on the extracted most occurrence and recent item sets from the huge data sets. The current traditional techniques can't provide the solution for all circumstances. The limitations are identified as need of dimension reduction methods and over fitting problems.

REFERENCES

- [1] MCChen, ALChiu, HHChang, "Mining changes in customer behavior in retail marketing", *Expert Syst. with Appl.* Vol28, Iss4, May-05, P.773-781.
- [2] G.Ertek, A.Demiriz, "A Framework for Visualizing Association Mining Results", Nov-2006, DOI:10.1109/11902140_63
- [3] MaryamNafari, Jamal Shahrabi "A temporal data mining approach for shelf-space allocation with consideration of product price", *Expert Syst. with Appl.*, Vol- 37, Iss-6, June-10, P.4066-72
- [4] Mutlu YA, "AR in DM: An App. on a Clothing and Accessory Specialty Store", *Canadian Soc.Sci.*, April-2014, DOI:10.3968/4541
- [5] WB Zulfikar, A Wahana, W Uriawan, N Lukman, "Implementation of association rules with a prior algorithm for increasing the quality of promotion", 2016 4th Intl Conf. Cyber and IT Service Mgmt, IEEE Xplorer, DOI: 10.1109/CITSM.2016/7577586
- [6] Abusida, "An association predict model: GECOL-Casestudy", *Intl.Jrnl.ofITandComp. Scie.*, 2019, doi:10.5815/ijitcs.2019.10.15
- [7] YNISari, ATiayudi, IraDS, "Impl.ofDM of predict food sales rate method using Apriori", *Intl.Jrnl.ofComp.Appl.*, July-2019, Vol-178No35
- [8] Sutisnawati, Reski, "Looking for transaction data pattern using Apriori and AR", *IOP conf.ser.:MatrSciand Eng.*, doi:10.1088/1757-899X/663/2/22078
- [9] Ariestya, W.W., Supriyatin, W. and Astuti, I. (2019) "Marketing Strategy for the Determination of Staple Consumer Products Using Fp-Growth and Apriori Algorithm", *Jurnal Ilmiah Ekonomi Bisnis*, 24(3), pp.225-235. doi:10.35760/eb.2019.v24i3.2229.
- [10] Anggraeni, S., Iha, M.A., Erawati, W., & Khairunnas, S. (2019). "Analysis of Sales by Using Apriori and FP Growth at PT. Panca Putra Solusindo". *Riset Dan E-Jurnal Manajemen Informatika Komputer*, 3(2), 41-46
- [11] Matobobo, C. and Osunmakinde, I.O. (2014) "A Comparative Market Basket Analysis For Centralized Retail Enterprises Using The ANN and AR Models". *C.CONTEMPORARY MANAGEMENT IN Proceedings of the 26th Annual SAIMS Conference*
- [12] Zeki Sušac, M. and Has, A. (2015) "Data Mining as Support to Knowledge Management in Marketing", *Business Systems Research Journal Systems Research Journal*, 6(2), pp. 186(2), doi:10.1515/bsrj-2015 -

0008.0008.0008.0008.0008.

[13] Matobobo Courage (2016) "Development Of An Intelligent Analytics Baced Model For Product Sales Opti. In Retail Enterprises" at the UNIVERSITY OF SOUTH AFRICA SUPERVISOR : PROFESSOR ISAAC O. SUNMAK INDE03July2016', (July).

[14] Y.J.Wang, QinXin,F.Coenen,"A Novel Rule Weighting Approach in Classification Association Rule Mining,"ICDMWorkshops 2007,Seventh IEEE International Conferenceon Data Mining Workshops,Pp.271–276,2007.