Middle-East Journal of Scientific Research 23 (1): 88-93, 2015
ISSN 1990-9233

DOI: 10.5829/idosi.mejsr.2015.23.01.22038

Designing an Expert System Using Association Rule Mining for Instant Business Intelligence

[1]Vijay Kumar Ravi, [1]P.M. Rahul and [2]S. Krishna Anand

[1]B. Tech Computer Science and Engineering, SASTRA University, Thanjavur, India 613 40
[2]School of Computing (CSE), SASTRA University, Thanjavur, India 613 401

Abstract: Over the past few years, Market Basket Analysis has been put into extensive use. The items most likely to be purchased can be generated using this analysis. Association rule mining is a very effective way for Market Basket Analysis by finding out the confidence of various products. With cost being one primary factor to be taken into consideration, it becomes imperatively necessary to decide suitable set of items incorporating the same. Besides, a learning component can also be developed for suggestion of offers. The aim is to develop an expert system that is very flexible and scalable to apply instant Business Intelligence for improving the profit. The same system could be extended to arenas such as stock markets and realtors.

Key words: Association Rule Mining · Learning · Market Basket Analysis · Price Sensitivity

INTRODUCTION

Association rule mining has been a successful method to perform market basket analysis. The confidence percentage generated through the association rule mining can be very effectively used to formulate the buying pattern of a group of people in an area. Scaling the number of parameters, the system can classify the results from the market basket analysis based on certain conditions. This classification will enable the seller to make a decision. It focuses on a particular scenario - buying pattern in departmental stores. There are three tasks involved. First of all, the task of determining the buying pattern based on the customer buying pattern is carried out. Secondly, the task of setting the price of the product to give a discount based on the customer buying preference is performed. Finally, the decision maker who actually decides which products to combine and sell in order to ensure the choice of that particular offer.

Market basket analysis is a prerequisite for the system. The generation of association rules is the main function where the confidence is generated based on the given input data and support count. The output data in the form of association rules is used by the decision making system. Determination of price component provides the final price of the product based on certain conditions. This output is also taken by the decision making system. Integration of the data mining and price determination is done by the decision making system. Use of expert system- learning component is necessary so that it evolves over time. The learning component evaluates the scenario based on real time parameters before deciding upon a particular decision.

Literature Survey: Data Mining techniques have been in use ever since 1900. Different mining techniques have been employed in a wide variety of applications. A significant contribution to the association rule mining was in the year 1993, when Rakesh Agarwal, Tomasz Imieliñsk and Arun Swami devised a formal model to mine association rules between sets of items from a large transactional database. They introduced a model based on the support count and the confidence percentage. Two parameters namely minimum support and minimum confidence were taken into consideration for the determination of association rules. In 1994, a fast algorithm for mining association rules, Hybrid Apriori algorithm, was devised by Rakesh Agarwal and Ramakrishnan. This paper focused on generating the confidence percentage using the candidate set generation. In 1997, cultural algorithms to support re-engineering of rule based expert systems was designed by Sternberg, M and Reynolds, R.G [1]. This paper dealt with

Corresponding Author: S. Krishna Anand, School of Computing (CSE), SASTRA University, Thanjavur, India 613 401.

fraud detection in a dynamic environment. Artificial Intelligence has been used in the field of medical diagnosis. In the year 2000, Kovalerchuk, Vityaev and Ruiz worked on the area of knowledge discovery in medical diagnosis and published their findings [2]. In the same year, another algorithm “FP Growth algorithm” was proposed by Han [3] which generated the support and confidence along with the association rules without the generation of candidate sets. Unlike Apriori algorithm, FP Growth algorithm is scalable and efficient for mining patterns. In 2001, a system for drop test analysis using intelligent data mining system was developed by Zhouv Nelson, Weimin Xiao, Tirpak and Lane [4]. In the same year, a paper dealing with building customer profiles using data mining methods was published by Adomaviciu, Tuzhilin[5]. In 2002, Aggarwal, Procopiuc and Yu formulated a model for finding localized associations in market data [6]. Qin Ding, Qiang Ding, Perrizo devised an efficient algorithm PRAM to mine rules from spatial data [7] in the year 2008. In 2010, an expert System was developed for Data Identification in Auction System. This model was integrated with Market Trends Analysis by Radoslav Fasuga, Tomas Drabek, Galina Toporkova, Libor Holu. In the same year, techniques were developed in automated industrial plants using data mining techniques and expert systems with learning capabilities by Alexandre Acácio De Andrade, Sergio Luiz Pereira, Eduardo Mario Dias and Caio Fernando Fontana [8].

Currently, the market basket analysis system completely depends on association rules to determine the customer buying pattern. The main advantage in the proposed system is that it does not completely rely on association rule mining for market basket analysis. Additionally, this system proposes an expert system which helps in decision making. Moreover, this system also integrates the price sensitivity concepts to help in predicting offers for the customers [9].

Proposed Model: The system designed focuses on a large number of aspects. However, three chief areas have been given enough emphasis. One of them deals with mining of association rules. The second part deals with price sensitivity based on the buying pattern. The final part integrates the first and second with the help of a learning centred expert system.

Data Mining: In data mining, association rules are mined from the given data set. The data set used here is the transactional data taken from a super market.

Table 1: Support Count and Confidence Percentage of Sample Data Set

Sl No	Premises	Conclusion	Support	Confidence
1	Meat, Butter	Bread	0.09465	0.403509
2	Butter	Bread	0.201646	0.404959
4	Milk, Coffee	Bread	0.119342	0.408451
5	Meat, Milk	Bread	0.074074	0.409091
6	Milk, Butter	Coffee	0.057613	0.424242
7	Meat, Coffee	Milk	0.098765	0.428571
8	Meat, Coffee	Bread	0.098765	0.428571
9	Milk	Bread	0.222222	0.442623
10	Coffee	Bread	0.222222	0.442623
11	Coffee, Bread	Meat	0.098765	0.444444
12	Bread	Milk	0.222222	0.446281
13	Bread	Coffee	0.222222	0.446281
14	Meat	Coffee	0.230453	0.448
15	Meat	Butter	0.234568	0.456
16	Coffee	Meat	0.230453	0.459016
17	Butter, Bread	Meat	0.09465	0.469388
18	Butter	Meat	0.234568	0.471074
19	Meat	Bread	0.259259	0.504
20	Bread	Meat	0.259259	0.520661
21	Milk, Bread	Coffee	0.119342	0.537037
23	Meat, Milk	Coffee	0.098765	0.545455
24	Milk	Coffee	0.292181	0.581967
25	Coffee	Milk	0.292181	0.581967

The association rules are generated using a standard algorithm such as Apriori or FP Growth. FP Growth algorithm is used for generation of the association rules in the designed system.

The two main parameters used in the mining of data sets are support count and confidence percentage.

Support Count deals with the determination of the probability of an item set occurring in the form of a transaction in a particular data set. In the system designed, the support count (Milk, Coffee, Bread) is 0.1193. This has been clearly highlighted in Table - 1 [10].

Confidence is defined as conf (x ◎ y) = Supp (x $\cup$ y) /Supp(x). One specific instance can be taken into consideration. Considering the confidence percentage for (Milk, Coffee, Bread), the confidence % has been computed and found to be equal to 40.8%.This has been clearly illustrated in the fourth row of Table - 1.

As seen in Fig. 1, the association rules mined and their corresponding support counts and confidence percentage have been calculated. Table 1 provides the exhaustive list of all combinations.

The values of four factors namely the premise and conclusion along with support and confidence have in turn been used by the expert system in framing the decision on the right kind of offer.

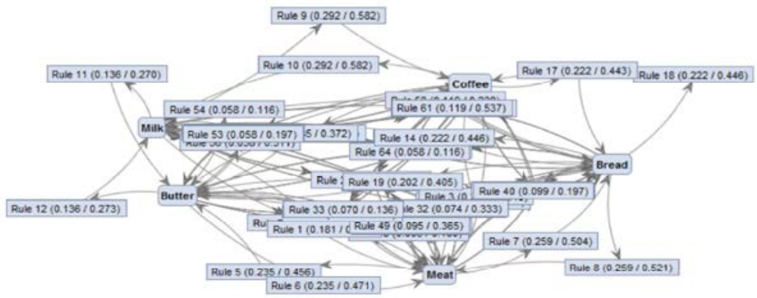

Fig. 1: Represents the mined association rules in the form of a graph.

Expert System Architecture:

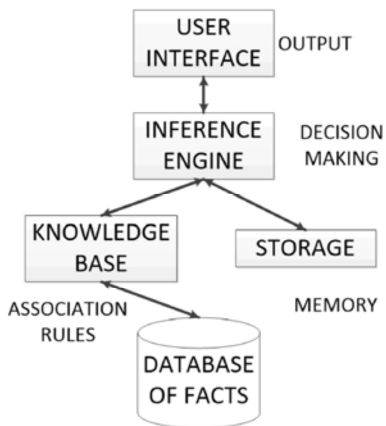

Fig. 2: Represents the Expert system model for the proposed system

Knowledge Base: Knowledge Base is the set of facts and rules that help the system in making a decision. These rules and facts are typically stored in a database. In the designed system, the transactional data from the super market is the database of facts for the system. From the database of facts, the association rules are generated. These association rules constitute the knowledge base of the system.

User Interface: User Interface is the domain where the human - machine interaction is made possible. The interface is designed in such a way that it takes the initial parameters of the system. These initial parameters are important for the calibration of the system. This forms the top layer in the architecture.

Memory Component: Storage/Memory is necessary when the system is running. The set of rules framed are stored in memory. This in turn could be used by the inference engine while framing conclusions.

Inference Engine: This is a vital component in an expert system. The inference engine takes the knowledge in the form of Association Rules and forms conclusions. In the designed system, combination of products with offers is found by the system. This in turn determines the level of utility.

Diagnosis of Expert System: The designed expert system can provide the most suitable offer for every category of customer. The special feature incorporated in this system is its ability to handle conflicts. Conflicts arise when the user belongs to more than one category. For instance, a student can be a part time employee also. In such a case, the two conflicting choice of offers are provided by the conventional system. This designed system goes a step further in trying to identify a choice of offers for that category also.

This system is also different for the conventional system in one more way. In conventional market basket analysis systems, association rules have to be generated from time to time when a set of transactions take place. The designed system generates the association rule only once. From the set of generated rules, the learning component learns the customer buying pattern.

Pre-Conditions: There are some prerequisite conditions that need to be established at the beginning of the run. These pre conditions determine the flow of program in the system. Based upon the conditions given, the expert system will decide on which offer to propose to the customer. Some of the conditions that are taken as a part of implementation are, “Age Group, Gender, Employment status, Time of the year, Time of the month”.

Age Group parameters consists of three sub groups namely, “Below 15, 15-30 and Above 30”. The product preference of the customer depends on these parameters. Employment Status parameter will have sub groups such

Table 2: Sample Database for Student Classification

Transaction Number	Items
T1	Pen, Paper
T2	Pen, Ink
T3	Pencil, Eraser, Paper
T4	Graph Sheet, Compass,
T5	Pencil, Paper, Eraser

as "Employed, Unemployed and Student". Similarly, the product preference also might depend upon the time of the year and time of the month parameters.

A student of age group in the range of fifteen to thirty has been taken into consideration, during the period of summer. The information has been mined and his/her product preference has been computed. Results show the chances of purchase of Stationeries and Cool Drinks are high. The corresponding function in propositional logic is shown below:

- if(agegroup is 15-30 and employment-status is student and summer)then buys(stationery,colddrink)

This is due to the fact that a student will mostly buy a stationery item and since it is summer he will prefer a cold drink rather than a hot drink.

It has also been observed that purchase of provisions by female users are high during the beginning and end of the month. The function in propositional logic has been shown below

- if(gender is Female and time of the month is Beginning or End) then buys(provisions)
- This rule described does not take the factor of Age into consideration.
- Similar computations have been performed for other user categories with high degrees of success.

Database: The real problem lies in the classification of the data. The products have to be classified prior to the execution of the system. The classification must be based on the parameters chosen. For instance, different groups of people may prefer different sets of products. This information can be obtained from a survey which will determine buying pattern of different groups of people. Upon completion of the classification process, the database will contain all the transactions in a particular classification.

Table 2 represents the sample database for a student who deals with only stationery items that students usually buy. In a real time scenario, the database might contain more transactions with products such as food items, uniform, shoes, bags and electronic gadgets. Similarly, in the same way, classification can also be done using other parameters.

Learning: Learning deals with updation of the knowledgebase from time to time. In the system considered, the likes and dislikes of the customers vary frequently. Besides, a wide variety of customers exist with varying degrees of likes and dislikes. The system designed incorporates a mechanism to identify the preferred set of items and providing suitable offer for the same.

Types of Learning: The ways of Learning are abundant. However, a suitable method appropriate to the type of problem must be chosen. In this connection, different types of learning have been analysed.

Supervised Learning: The algorithm maps inputs to outputs based on a function. The system is trained on labelled examples. Some of the algorithms used are Analytical Learning, Back propagation and Bayesian Classification. Different applications where supervised learning are used include Bioinformatics, Database Marketing and Handwriting Recognition.

Unsupervised Learning: Self-organizing map (SOM) and adaptive resonance theory (ART) are commonly used algorithms in Unsupervised Learning. Unsupervised learning is widely used to reduce high dimensional genomic data involving huge number of sets.

Semi-Supervised Learning: Semi supervised learning combines both labelled and unlabelled examples and an appropriate classifier is generated. It is a type of supervised learning mechanism. There are several assumptions that are used in semi supervised learning such as Smoothness Assumption, Cluster Assumption and Manifold Assumption.

Reinforcement Learning: The system observes a scenario and then a policy is derived based on that scenario. The learning algorithm is guided by the feedback provided for every action having some impact in the environment.

Transduction: It is similar to supervised learning, but without explicitly constructing a function, the algorithm tries to predict new outputs. The parameters used to redict are the training inputs, new inputs and training outputs. The three commonly used algorithms for transduction are Partitioning Transduction, Agglomerative Transduction and Manifold Transduction.

Learning to Learn: Algorithm learns on its own based on the previous experience. In this type of learning, the system can increase its accuracy through learning by experience. Two prominent applications in which this type of learning is used are Spam filtering and Web Search

The technique of learning by explaining experiences has been incorporated in this work owing to the list of advantages cited. The main advantage of the learning based system is that it focuses on maximizing the utility of the system while achieving its goal and also acts rationally by learning. Learning enables the system to operate in an obscure environment. Besides, the level of competency is improved.

In the proposed system, by learning from previous experiences, it is possible to suggest offers based on the given parameters such as age group and employment status. This is achieved by having a utility parameter. A utility parameter is fixed to each and every offer generated by association rule mining. Then if that particular offer is chosen by the customer, then the utility value of that particular offer increases. Another parameter in addition to the utility parameter namely timestamp has been taken into consideration during the process of learning. This deals with recording the period during which the offer was preferred by the customer. This value is in turn stored in the database.

If (Milk, Bread, Coffee) is chosen by the customer, then the utility value of the corresponding offer increases. The utility value is a simple counter value which keeps track of the number of times the offer has been chosen. A sample pseudo code for the same has been shown below.

- if (buys (offer1)) then
- offer1.counter = offer1.counter+1
- offer1.time_stamp= current_time

In the pseudocode shown, Offer 1 corresponds to the combination of Milk, Bread and Coffee. The expert system evaluates all the possible offers based on the utility parameters namely counter and time_stamp. The parameter time_stamp is used when there are two or more offers that have the same counter value.

The pseudocode shown below uses the secondary utility value namely time_stamp on deciding which offer is to be displayed when the counter value of two offers have same counter value. This can be extended to situations where there are more than two offers to evaluate.

- if (offer1.counter>offer2.counter)) then display offer 1
- else if(offer1.counter<offer2.counter)then
- display offer2
- else then
- if(offer1.time_stamp>offer2.time_stamp) then
- display offer1
- else
- display offer 2
- end if
- end if

Price Sensitivity: Estimating the price of the product depend on a wide variety of parameters. The method of estimation of price is very necessary in the retail industry. The estimation may be based on customer buying behaviour for a particular product. The main factor to be considered here while setting the price for a particular product is that once the price is set, it should not make the customer to stop buying it. This can be made sure by fixing the elasticities in the price of the product. This price elasticity will indicate the consumer sensitivity to various changes in price that are located around a specific reference price. This elasticity in price is referred to as latitude of acceptance [9]. The latitude of acceptance determines the gain and loss threshold based on certain conditions.

In the designed system, the price of the product will be determined by the customer buying behaviour. The expert system decides the combination of the offers and the price discount that can be provided. The system uses the utility parameter when deciding the discount that can be provided to the particular offer. In the designed system, the utility parameters such as counter value and time_stamp are used as parameters for deciding the discount percentage. As the utility parameter is used to keep track of the customer buying preference, it is easy to provide discount based on the utility parameter.

As seen in Table 3, the counter value of the record with Id = 2 is 8 which is maximum. So this implies that the customer prefers this product more. Hence, it is given the maximum discount. Similarly, the discount values can be set to all the offers. The main advantage of this model is that the discount amount can be dynamically changed in accordance with the customer buying pattern.

Table 3: Database Values Showing Counter and Time_stamp

LT15BOYSSTUDENTS						
Id	Premise	Conclusion	Support	Confidence	Ctr	Time_Stamp
1	Sharpner	Pencils	0.25	1	3	35:07.8
2	Ruler	Pencils	0.2	1	8	38:38.0
3	Mouthwash, Sharpner	Pencils	0.2	1	5	59:16.5
4	Pencils	Mouthwash	0.25	0.83	6	38:49.7
5	Mouthwash	Pencils	0.25	0.833333	4	59:28.2
6	Pencils	Sharpner	0.25	0.833333	3	35:23.3
7	Hair Gel	Antiperspirant / Deodorant	0.4	0.8	3	35:20.4

CONCLUSION

An intelligent rule based expert system has been designed to assist the customer in providing a suitable offer based on purchase history. A technique in association rule mining has been incorporated in the same. The system learns using the property of prior experience and automatically updates the knowledge base. Besides, additional work has also been carried out. This work focuses on providing suitable discounts to offers based on the customer preferences.

This work could be further enhanced by designing a suitable neural network to find relationship between the customer and his/her preferences. Other forms of learning could also tried out at appropriate stages. However, despite the minor seemingly insignificant errors or omissions that would have been inadvertently committed, the designed model is found to be highly efficient in determining suitable offer that could be provided to the customers.

REFERENCES

1. Robert, G. Reynolds and Michael Sternberg, 1997. Using Cultural Algorithms to support Re-engineering of Rule-based Expert Systems in Dynamic Performance Environments: a case study in Fraud Detection, IEEE Transaction on Evolutionary Computation, USA, 1(4): 225-243.
2. Kovalerchuk, B., E. Vityaev and J.F. Ruiz, 2000. Consistent knowledge discovery in medical diagnosis, IEEE Eng. Med. Biol., 19: 26 -37.
3. Jiawei Han, Jian Pei, Yiwen Yin, 2000. Mining frequent patterns without candidate generation, proceedings of the 2000 ACM SIGMOD international conference on Management of data, Dallas, Texas, USA, 1-12: 15-18.
4. Zhou, C., P. C. Nelson, W. Xiao, T. M. Tirpak and S.A. Lane, 2001. An intelligent data mining system for drop test analysis of electronic products, IEEE Trans. Electron. Package Manuf., 24(3):222 -231.
5. Adomavicius, G. and A. Tuzhilin, 2001. Using data miningmethods to build customer profiles, IEEE Computer, 34(2): 74-82.
6. Aggarwal, C.C., C. Procopiuc and P.S. Yu, 2002. Finding Localized Associations in M1arket Basket Data, IEEE Trans. on Knowl. and Data Eng., 14(1): 51-62.
7. Ding, Q., Q. Ding and W. Perrizo, 2008. PARM-An efficient algorithm to mine association rules from spatial data, IEEE Trans. Syst. Man, Cybern. B, Cybern., 38(6): 1513 -1524.
8. Fasuga Radoslav, Tomas Drabek, Galina Toporkova and Libor Holub, 2010. Expert System for Data Identification in Auction System along with Market Trends Analysis, IEEE 7th International Conference on e-Business Engineering.
9. Rakesh Agrawal and Ramakrishnan Srikant, 1994. Fast Algorithms for Mining Association Rules in Large Databases, Proceedings of the 20th International Conference on Very Large Data Bases, 12-15: 487-499.
10. Casadoa Esteban and Juan-Carlos Ferrer, 2013. Consumer price sensitivity in the retail industry: Latitude of acceptance with heterogeneous demand, European Journal of Operational Research, 228(2): 418-426.