DATA SCIENCE AND KNOWLEDGE DISCOVERY THROUGH DATA MINING PARADIGMS

1Indu Chhabra, 2Gunmala Suri

1 Professor, Department of Computer Science and Applications, Panjab University, Chandigarh
2Professor, University business School, Panjab University, Chandigarh

Email: 1 indu_c@pu.ac.in, chhabra_j@rediffmail.com, 2 g_suri@pu.ac.in

https://doi.org/10.26782/jmcms.2019.04.00014

Abstract

Current trends in software development have shown a strong move towards autonomous and rational mechanism for the human societal growth. Customer behavior analysis and its knowledge have always been given its due importance in research community to develop real life practical solutions. In this scenario a real-world phenomenon of customer buying habits is tested through observations lying in the database and is experimented and validated through association mining.

On the flip side of the coin, the development of intellectual and evolutionary data mining tool for retail industries through the machine learning algorithm has always been proved to adequately respond to environment changes and improve its behavioral rules to derive intelligent quotient. A case study of Market basket analysis is simulated to imitate customer behavior in the dynamic environment to predict about rational and intelligent behavior for future business expansion.

Keywords: Customer behavior analysis, Data mining, Intellectual Management, Neural Networks, Genetic algorithm, Retail industry

I. Introduction

Growing business segments like e-commerce, e-banking, social networking and mobile applications has pushed the need for more and more creative technologies to extract business intelligence for societal development. The ever growing data mining technology is getting its due importance because of its potential in solving versatile problems as it provides a platform to link common people to product providers.

The main function of data mining technology is generally divided into five kinds including classification, estimate, clustering, association rule and prediction. Association
rule mining is an active research area with numerous algorithms used for knowledge discovery. The road map of the research problem is:

- Problem Recognition
- Related facts and figures search for system inputs
- Evaluation of genetic tool
- Purchase analysis and implementation
- Post purchase behavior evaluation for Knowledge discovery

The present research provides the practical perspective for the investigation of underlying problem and its usefulness and relevance. For cooperative and collaborative relationship between buyers and sellers, strategies and processes it should promise to individualize and personalize the vital customer information at every point of customer interface.

The current work is focused on analyzing customer behavior for most frequent items to be purchased together to derive most interesting verified market facts hence to execute the most appropriate market laws and rules for the overall business growth.

II. Problem Statement

Data mining is applied for the automatic selection of most interesting facts to provide business benefits. The underlying research problem is an attempt with an objective of providing intelligent solution to the problem of analyzing customer buying habits through association rule mining algorithm of Market Basket Analysis of the paired itemsets commonly purchased by consumers. The database is based on human habits facts derived through varying verbal and non-verbal communication modes.

The introductory section gives details about the budding data mining technology. After illustrating the research problem, its application in fusion with intelligent association mining has been explored through the exhaustive literature survey for diversified areas. After describing the proposed methodology, to demonstrate, how these systems are helpful in providing real time solutions, a case study of “How data mining is a competitive source for consumer behavior analysis in retail industries to improve on productivity”, is analyzed through the real time practical calculation of controlling parameters of support, confidence to derive the knowledge about the of buying behaviour of a user to frame out the production and sales policies in reference to Market basket Analysis.

III. Review of Literature

This very promising data mining technology is emerging as a new way to emulate the corresponding conceptual paradigm for designing customized systems as well as for dealing with the complexity of human behavior [IV]. Numerous approaches exist, attempting to optimally reflect both the inner states as well as perceived environment of a
customer. For real world applications, the authors have studied an association mining algorithm for automatic extraction of association rules[IX]. The Ahmed[I] in his paper “Applications of data mining in retail business” has tried to improve data mining analysis by applying several methods including clustering, principal component analysis and discriminate analysis. S. Vijaylaxmi, V. Mohan, S. Suresh Raju in their paper “Mining of users’ access behavior for frequent sequential pattern from web logs” explain how sequential pattern mining comes from association rule mining[VIII]. As comprehensibility and predictive accuracy, the measures of the quality of the rule, can be considered as different objectives of the association rule mining problem so for association rule mining problem, are modeled as a multi-objective problem rather than as a single-objective problem [VII] for the development of applications taking the advantage of Multi-Store environment.

Customer data and information technology (IT) tools form the foundation upon which any successful CRM strategy is built. In addition, the rapid growth of the internet and its associated technologies has greatly increased the opportunities for marketing and has transformed the way of how the relationships between companies and their customers are managed. Nagi [V] has defined it as influencing customer behavior through meaningful communications in order to improve customer retention and customer profitability. As data mining is an attitude which suggests that the business actions should be based on learning. The informed decisions are better than the uninformed ones. The neural network tries to obtain the measuring results which are beneficial to the business [III]. Based on the extracted interesting patterns, few strong postulates in form of rules are derived to scale these quality rules into generalized quantitative properties of support and confidence [VI]. The support and confidence are the metrics which are utilized together to threshold the level of interestingness in the picked rules. Hence these parameters facilitate in finding those interesting patterns which can improve on comprehensibility and predictive accuracy of the system [II].

IV. Process Framework and Methodology

The "Market Basket Analysis", also known as association analysis is implemented for the better understanding of customer buying preferences. It addresses the various questions of “What goes with what”? Which items are ordered and purchased together?” For a given transaction database T, if an association rule $X \rightarrow Y$ holds then it has the evidence with confidence $\tau \%$ and $\sigma \%$ support for transaction set T to provide $XUY$ combined set. For association rule mining firstly frequent patterns with respect to minimum support threshold are mined. Then association rules are generated with respect to minimum confidence as indicated by confidence threshold.

It is the three phase process of Database design, Association Rule Discovery and Recommendations as shown in Fig 1.
Fig 1. Mining Scenario

Phase 1: In database design, the transactions of retail data are grouped together by product category to graph the item frequencies. Then the Indicator matrix is created to show the intensity and frequency of each combination. In the present analysis, 10 market basket transactions worth of point-of-sale retail data, where each basket represents a combination of items on a computer peripheral store receipt, are recorded for the further analysis.

Phase 2 Association rule discovery phase short lists the parameters to be analyzed further. It generates the correct baseline for developing quality combinations through the controlling parameters of support and confidence to verify them for the given different sets of transactions.

Phase 3 Recommendations

The knowledge inferred from the found frequent itemsets patterns is analyzed to recommend the most suitable combination for the general problem solving.

V. Performance Efficiency Assessment:

The objective is to design an algorithm for the classification and analysis of the discovered interested facts through data mining. Hence the focus of this research is to evaluate the algorithm in such a way that they provide frequent profitable patterns to help market analyst to frame the best informed decisions for improving their business with the consideration of future risks.

Phase 1: Generating Database through Indicator Matrix: where rows are the transactions of the basket of items purchased together by a single customer and columns are the items itself with
\( A_{ij} \): Customer i purchased item \( j \).

| Basket-ID | A | B | C | D | E |
|-----------|---|---|---|---|---|
| T_1       | 1 |   |   |   | 1 |
| T_2       | 1 | 1 | 1 |   |   |
| T_3       | 1 |   | 1 |   |   |
| T_4       |   | 1 | 1 | 1 |   |
| T_5       | 1 | 1 | 1 | 1 |   |
| T_6       | 1 |   |   | 1 |   |
| T_7       | 1 |   | 1 | 1 |   |
| T_8       | 1 |   | 1 |   | 1 |
| T_9       | 1 | 1 |   | 1 |   |
| T_{10}    |   | 1 | 1 |   |   |

Phase 2: Association Rule Mining:

Step 1: If \((A_{i1} = 1) A (A_{i2} = 1) \ldots (A_{ik} = 1) \rightarrow A_{i(k+1)} = 1\) with \(1 \leq ij \leq p\) for all \(j\).

Step 2: Finding Association Rules through Support (Frequency) and Accuracy (Confidence) calculation where \(\theta\) = Itemsets and \(fr(\theta)\) = No. of cases which satisfy \(\theta\). Hence

\[
fr(\theta A \psi) = \text{Support} = \text{No. of cases of } \psi \text{ in the database those satisfy } \theta.
\]

So rule is \(C ((\theta A \psi)) = \frac{fr(\theta A \psi)}{fr(\theta)}\)

Step 3: Finding frequent Patterns: \{A\}, \{B\}, \{C\}, \{D\}, \{AC\}, \{BC\} having number of ones 4 or greater in their respective columns.

Step 4: Knowledge Discovery:

\[
\text{Rule } (A \rightarrow C) = \frac{fr(A \land C)}{fr(A)} = \frac{\text{No. of cases of } C \text{ in the database those satisfy } A \text{ also}}{\text{No. of ones in } A} = \frac{4}{6} = 0.66
\]

\[
\text{Rule } (B \rightarrow C) = \frac{fr(B \land C)}{fr(B)} = \frac{\text{No. of ones in both } B \text{ and } C}{\text{No. of ones in } B} = \frac{5}{5} = 1 \quad \text{... *}
\]

As the minimum confidence is 0.7 and as for rule \((B \rightarrow C)\) \(1 > 0.7\) so it is sure that B and C items will be purchased together.

Phase 3: Recommendation:

For \(A=\) Pen drive, \(B=\) Antivirus, \(C=\) Laptop, \(D=\) DVD and \(E=\) Digital camera, the generalization about customer behavior is:
66% chances of buying Laptop along with Pen drive

100% chances of buying Laptop + Antivirus

VI. Research Contribution

In the present context of behavior analysis, the evaluated data mining is capable of converting the psychology of consumers into statistical format to justify how he reasons and selects between different alternatives. Hence the association of the customers with different products is converted into a technical format by which one can analyze the buying behavior to verify the most appropriate market laws and rules.

VII. Future Scope:

This algorithms is little costly, in case of large number of patterns because in that case the pruning technique is insufficient to help market analyst to make decisions such as planning the supermarket shelf space, changing the previous layout, setting new store layouts and new product assortment to verify the group of products to be put on promotion for marketing profits. Rare item sets provide useful information in different decision-making domains such as medical, security, fraudulent transactions and retail communities. For example, in a supermarket, customers purchase microwave ovens or frying pans rarely as compared to bread, washing powder, soap. But the former transactions yield more profit for the supermarket.

VIII. Conclusion

The present study is an effort towards the demonstration of data mining application on the voluminous data through the case study of Market basket analysis. The filtration of most relevant facts is analyzed through the intellectual paradigm of association mining. The outputs are the rules, to be transformed to the corresponding business policies for the trade expansion. Hence the management may update their system of providing various types of services to their customers at one stop to delight and retain the customer with the same business group and retailer. With this study the researchers can conclude about the previously unknown and potentially useful information from vast resources of raw data.
I. Ahmed, S.R., “Applications of data mining in retail business”, Information Technology: Coding and Computing, vol. 2, 2004, pp. 455-459.

II. Amandeep Kaur, P. S. Grover, “Performance Efficiency Assessment for Software Systems” a chapter in “Advances in Intelligent Systems and Computing” book series, AISC, Volume 731, June 2018.

III. Ansari Azarnoush and Riasi Arash, “Customer Clustering Using a Combination of Fuzzy C-Means and Genetic Algorithms”, International Journal of Business and Management; vol. 11, Canadian Center of Science and Education, June 2016, ISSN 1833-3850.

IV. Ismail, J., “The design of an e-Learning System: Beyond the hype”, Internet and Higher Education, vol. 4, 2002.

V. Ngai, E.W.T, Xiu, Li and Chau, D.C.K., “Application of data mining techniques in customer relationship management: A literature review and classification”, Expert systems with applications, vol. 36, March 2009, pp. 2592-2602.

VI. Pillai Jyothi, “User centric approach to itemset utility mining in Market Basket Analysis”, International Journal on Computer Science and Engineering, Jan 2011.

VII. Sandhu Parvinder, Dalvinder S. and Panda S. N, “Mining utility-oriented association rules: An efficient approach based on profit and quantity”, International Journal of the Physical Sciences, vol. 6, pp. 301-307, Jan 2011.

VIII. Vijayalakshmi S., Mohan V., Suresh Raja S., “Mining of users’ access behavior for frequent sequential pattern from web logs”, International Journal of Database Management System (IJDM), vol. 2, August 2010.

IX. Woo, J.Y., Bae, S.M., and Park, S.C., “Visualization method for customer targeting using customer map”, Expert Systems with Applications, vol. 28, 2005, pp. 763-772.