

Enhancing The Association Rule Mining Technique

Er. Naina Choukade¹, Er. Ketan Khandelwal²

Department of CSE

¹M.Tec Scholar, Sushila Devi Bansal College of Technology (SDBCT), Indore

² Assistant Professor, Sushila Devi Bansal College of Technology (SDBCT), Indore

Abstract- *The term association can be used for combination of objects. Additionally the association rule mining provides us information about the association and their combinations with other objects. Therefore this technique is very useful for various real world applications such as security, prediction, recommendations and others. In this paper we are providing a survey on recently developed technique that is used for association rule mining and the contributions that are claimed to enhance the performance of traditional association rule mining. In addition of that the paper includes a new technique that helps to select the optimal rules generated from the different association rule mining techniques. Finally the paper is concluded and future extension of the given work is described in this paper.*

Keywords- association rule mining, survey, rule mining techniques, association rule mining, and optimization of rule selection.

I. INTRODUCTION

The association rule mining is a classical domain of data mining [1]. In data mining techniques classification, prediction, clustering and association rule mining are frequently used algorithms. Among association rule mining is different from other techniques. That technique can be used for classification, prediction and other decision making tasks [2]. Therefore the association rule mining is an essential technique of data mining [3]. The data mining technique supports two different kinds of modeling first the opaque data models and seconds the transparent data models. The opaque data models are the techniques that analyze data and direct generates outcome for their applications. There is not any provision to know how the values are produced. On the other hand in transparent data modeling the rules or tree or any kinds of matrix is generated by which the computation of the outcomes become transparent and we can estimate the outcome by paper and pencil. In this context the association rule mining technique is a transparent data modeling technique for mining information from raw or transactional data [4]. In this presented work the association rules mining techniques are the key area of study and application design.

The main aim of the proposed work is to study and design an improved technique of data mining by which the

optimal rules from the different algorithms can be selected. Therefore the proposed work first investigates about the different techniques such as apriori and FP-tree algorithms [5] which are used to generate the association rule mining. In next the efforts are made to optimize the rules for finding the most optimal rules from the existing rules or generated rules from the other algorithms. Therefore this paper is focused on searching the recently developed approaches for generating or optimizing the association rules. This section provides the overview of the proposed work involved in this paper. In further sections the survey of existing methodologies and the proposed data model for association rule mining is involved.

II. LITERATURE SURVEY

This section involves the different research efforts and the techniques that are contributing in association rule mining or optimizing the techniques of rule selection.

Market Basket Analysis(MBA) otherwise called association rule learning or liking examination, is an information mining strategy that can be utilized in different fields, for example, showcasing, bioinformatics, instruction field, atomic science and so forth. The fundamental point of MBA in advertising is to give the data to the retailer to comprehend the buy conduct of the purchaser, which can help the retailer in adjust basic leadership. There are different calculations are accessible for performing MBA. The current calculations take a shot at static information and they don't catch changes in information with time. In any case, proposed calculation mine static information as well as gives another approach to consider changes occurring in information. Manpreet Kaur et al [6] talks about the information mining procedure i.e. association rule mining and give another calculation which may accommodating to look at the client conduct and helps with expanding the deals.

Street mischance is one of the essential territories of research in India. An assortment of research has been done on information gathered through police records covering a restricted segment of parkways. The examination of such information can just uncover data with respect to that part just; however mishaps are scattered on parkways as well as on neighborhood streets. An alternate wellspring of street mishap information in India is Emergency Management examine

Institute (EMRI) which serves and monitors each mischance record on each sort of street and cover data of whole State's street mishaps. In this paper, Sachin Kumar et al [7] have utilized information mining strategies to examine the information given by EMRI in which we first group the mishap information and further association rule mining system is connected to recognize conditions in which a mischance may happen for each bunch. The outcomes can be used to put some mischance aversion endeavors in the zones distinguished for various classifications of mishaps to conquer the quantity of mischances.

In this paper, Ömer M. Soysal et al [8] address the issue of mining organized information to discover possibly helpful examples by association rule mining. Unique in relation to the customary locate all-then-prune approach, a heuristic technique is proposed to remove generally related examples (MASPs). This approach uses a maximally association requirement to produce designs without looking through the whole grid of thing mixes. This approach does not require a pruning procedure. The proposed approach requires less computational assets as far as time and memory necessities while producing a long arrangement of examples that have the most noteworthy co-event. Besides, k-thing examples can be gotten on account of the sub-grid property of the MASPs. Likewise, the calculation delivers a tree of the identified examples; this tree can help chiefs for visual examination of information. The result of the calculation executed is outlined utilizing auto collision information. The proposed approach can possibly be used in huge information examination.

Recognizing new item openings must be an essential for an association's maintainable development since it can help make new market portions. In such manner, various investigations have endeavored to recommend precise techniques to find new mechanical chances. Be that as it may, from these techniques, it is hard to make sense of which items can come into the market because of the mechanical chances. In addition, they have attempted to gauge non specific potential qualities without considering a particular target firm so it is difficult to judge whether the found open doors are in fact achievable to the objective firm. These issues have a tendency to diminish the common sense of the found mechanical chances. Along these lines, Wonchul Seo et al [9] proposes a deliberate way to deal with distinguish potential item openings by mirroring the objective association's inside abilities. The abilities are inalienably inconspicuous so we have to make sense of substitutes for the company's capacities. The current items having a place with a firm can be by and large a reason for growing new items. The firm is now great at managing the current items so we consider the association's current item

portfolios its inner abilities. We first concentrate item data from patent database utilizing content mining method, and after that create item association rules spoke to as coordinated sets of items. At long last, we assess potential estimation of item openings considering an association's inward capacities. An exact investigation is directed to demonstrate the relevance of the exhibited approach utilizing licenses allowed in the United States Patent and Trademark Office amid 2009 and 2013. We expect that our approach can encourage item arranged R&D by introducing a front-end show for new item advancement and determining attainable item openings as indicated by the objective company's interior capacities. In addition, the displayed efficient approach can be a reason for a R&D arranging framework that can help R&D organizers in performing item situated innovation arranging exercises.

One of the very much looked into and most critical systems of mining information is Association Rule Mining. Association Rules as the name itself shows incorporates discovering relationships among sets of things in exchange database. Most well known calculation of association rule mining is Apriori is utilized for information revelation. The proposed work by Er. Sukhvir Kaur et al [10] depends on discovering association rules considering the multidimensionality of the qualities and lessening the calculation time that will expand the productivity. Proposed work will enhance the current Apriori calculation and will lessen a portion of the downsides of the current calculation.

III. METHODOLOGY

Association rule mining technique is one of the most adoptable techniques in data mining technique. In a number of applications apriori, FP-Tree and éclat algorithms are used. Among them FP-Tree and Apriori algorithm is most popular algorithms. The apriori algorithm is sometimes computationally expensive because of generation of candidate sets. On the other hand the FP-Tree algorithm is efficient as compared to the apriori algorithm but sometimes it generates more than one tree for rule generation. In addition of that for the similar set of data both the algorithms can generate different amount of rules and different from each other. In this context a new data model suggested that help to optimize or select the most suitable rules for mining association rules. The figure 1 shows the proposed data model for optimal rule selection.

The proposed data model is demonstrated in the above given figure 1. In this diagram the input of the system is transactional data base. The transaction of a data consists of different number of item set. Additionally based on the item sets combinations the association rules are extracted from data.

After providing input to the system the dataset is processed for extraction of item set and their frequency in dataset. After that obtained information from dataset is utilized with the apriori algorithm and the FP-Tree. Both the algorithm processes the dataset and generates the association rules according to their own processes. After generation of rules from both the data models the rules are collected for optimization. In order to minimize the amount of rules and for finding the most appropriate rules from the generated set of rules. The genetic algorithm is employed on these rules. First the algorithm selects those rules are common in both the rule set and then the optimization process using genetic processes is obtained. After completing the genetic process the remaining rules are optimal set of rules.

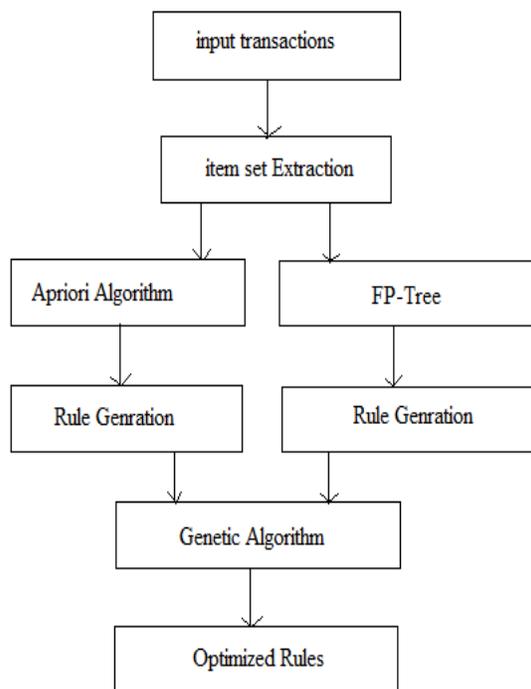


Figure 1 Methodology

This section provides the basic concept of the proposed model. The conclusion of conducted work and the future extension off the work is described in next section.

IV. CONCLUSION

The association rules are used in a number of applications for decision making, prediction and security purpose. Therefore the association rule mining is an essential domain of data mining and data analytics. The proposed work is focused on study of association rule mining techniques. Therefore different research contributions about the association rule mining technique is included in this paper. Finally a data mining model is proposed that can optimize the rules generated from the different techniques of association rule generation

techniques. In this context the genetic algorithm is proposed for optimizing the association rules. in near future the proposed data model implemented and their performance is demonstrated.

REFERENCES

- [1] Bhandari, Akshita, Ashutosh Gupta, and Debasis Das, "Improvise apriori algorithm using frequent pattern tree for real time applications in data mining", *Procedia Computer Science* 46 (2015): 644-651.
- [2] Rana Ishita and Rana Ishita, "Frequent Itemset Mining in Data Mining: A Survey", *International Journal of Computer Applications (IJCA)*, Volume 139 – No.9, April 2016
- [3] Sanjaydeep Singh Lodhi and Premnarayan Arya, "Frequent Itemset Mining Technique in Data Mining", *International Journal of Advanced Research in Computer Engineering & Technology* Volume 1, Issue 5, July 2012
- [4] Borgelt, Christian. "Frequent item set mining." *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* 2.6 (2012): pp. 437-456.
- [5] Goethals, Bart. "Frequent set mining." *Data mining and knowledge discovery handbook* (2005): pp. 377-397.
- [6] Manpreet Kaur, Shivani Kanga, "Market Basket Analysis: Identify the changing trends of market data using association rule mining", *Procedia Computer Science* 85 (2016) 78 – 85
- [7] Sachin Kumar, Durga Toshniwal, "Analysing Road Accident Data Using Association Rule Mining", 978-1-4673-9354-6/15/\$31.00 ©2015 IEEE
- [8] Ömer M. Soysal, "Association rule mining with mostly associated sequential patterns", *Expert Systems with Applications* 42 (2015) 2582–2592
- [9] Wonchul Seo, Janghyeok Yoon, Hyunseok Park, Byoung-youll Coh, Jae-Min Lee, Oh-Jin Kwon, "Product opportunity identification based on internal capabilities using text mining and association rule mining", *Technological Forecasting & Social Change* 105 (2016) 94–104
- [10] Er. Sukhvir Kaur, Er. Harjeet Singh, "An Enhanced Data Mining Technique to improve the CRM using Association Rule Mining", *International Journal of Engineering Science and Computing*, July 2018, Volume 8 Issue No.7