ISSN: 2321-7782 (Online)
Volume 4, Issue 1, January 2016
International Journal of Advance Research in Computer Science and Management Studies
Research Article / Survey Paper / Case Study
Available online at: www.ijarcsms.com

Multi-Dimensional Rule Hiding Algorithm for Transactional Database

Dr. K. Kavitha
Assistant Professor
Department of Computer Science
Mother Teresa Women's University, Kodaikanal, India

Abstract: ***Data mining is a rapidly growing technology for extracting valuable information based on decision making association role mining is an important application explored in data mining. The main objective of the research work is to identify the interesting rules. Many algorithms have been suggested over the last two decodes for market Basket analysis. This paper suggested Multidimensional Rule Hiding (MDRH) algorithm for generating the most effective interesting values that satisfies the constraint Weight.***

KEYWORD: Association Rule, Support, Confidence, Weight, Lift, Quantitative, Binary.

I. INTRODUCTION

Data mining is the process of discovering meaningful patterns and trends by shifting large amount of data stored in repositories, using pattern recognition technologies as well as statistical and mathematical techniques. Data Mining is a new technology and rapidly growing field which used in extracting valuable information from data warehouses and databases.

Association rule mining is an essential part of data mining technique due to the descriptive and easily understandable nature of the rules. Rakesh Agrawal.et.al introduced a common way of measuring the usefulness of association rules by support & confidence framework [1].

Organisation of the Research Work:

» Related work is presented in section II.

» Section III, proposed work is presented

» Section IV, Proposed algorithm is suggested

» Section V concludes research work.

II. RELATED WORK

2.1 Apriori Algorithm

Apriori algorithm is one of the classical algorithms in data mining technique, which is based on a candidate set generation logic. Apriori algorithm suffers repeated database scan in generating association rules. It generates candidate itemsets and tests whether it is frequent or not[2].

Steps:

1. Find the frequent itemsets, i.e., find all I with support ≥min_sup.{If I = (A,B) is frequent, then both A and B are also frequent}.

2. Find frequent itemsets with cardinality ranging from 1 to k.

3. Generate Association Rules from frequent itemsets.

The first pass of the algorithm simply counts item occurrences to determine the large 1itemsets.Subsequent pass, say pass k, consists of two phases. First, the large itemsets L_k-1 found in the $(k-1)^{th}$ pass are used to generate candidate itemsets C_k , using the apriori gen function[4]. Next, the database is scanned and the support of candidates in C_k is counted. For fast counting, to determine the candidates in Ck that are contained in a given transaction t efficiently.

2.2 FP-Growth Algorithm

The FP-Growth approach is based on divide and conquers strategy for producing the frequent item sets It is a two step approach to generate frequent itemsets without candidate generation. The algorithm is based upon a tree representation of frequent itemsets. It compresses a large database of transaction into a compact frequent –pattern tree structure[3].

In FP-tree scan the database once to collect all frequent items and their support counts. All frequent items are sorted in descending order of support and denoted as L. After that, create the root of an FP-tree and label it as "null". Scan the database for a second time. The items of each transaction in the database are sorted according to the order of *L*.

On inserting a transaction, if the tree has the same path, then the count of each node in the path increases if the path is incomplete in the tree, then a new branch and new nodes are created. After constructing the FP-tree, the FP-growth algorithm recursively builds a conditional pattern base and conditional FP-tree. Then, it generates all frequent patterns.

2.3 Multidimensional Quantitative Rule Generation Algorithm

This research work introduces a new multidimensional quantitative method to handle the categorical attributes and the numerical attributes in an efficient way. The conversion of data into quantitative method is taken place rapidly. Using the binary patterns, frequent patterns are identified using FPgrowth with minimum confidence threshold limit.. The conversion reveals all the frequent patterns from the database [6].

III. PROPOSED METHODOLOGY

The proposed method generates the dataset based on the attribute for the transactional database. Data conversion and concept hierarchy are performed to split up the attribute values according to the threshold limit. Multidimensional quantitative set frames the relationships and extract the effective association rule mining. The figure-1 represents the overview of the MDRF algorithm.

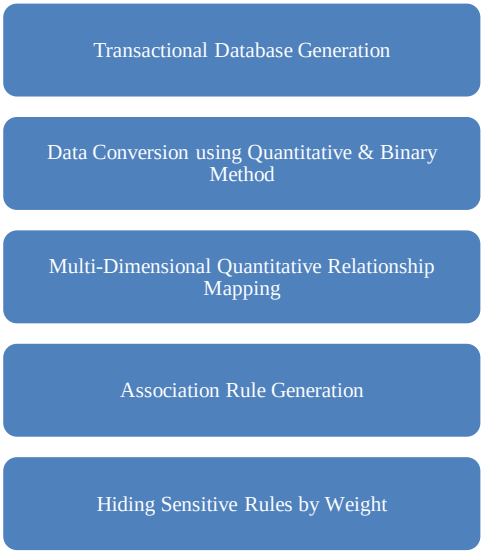

Figure-1 Overview of MDRH Algorithm

Database is a combination of both categorical and numerical values. First categorical attribute are converted into integer values which reduces the database scanning time. This conversion makes all the attributes in to numerical values. In this

approach, binary values are constructed according to the specified constraints related to the problem. For example, Annual_Income>=500000 is assigned as 1 and Annual_Income <500000 is assigned as 0. Similarly, categorical attributes are divided into two categories as per the hierarchical concept and assigned the number as 0 and 1[9]. The table 1 shows the sample dataset.

Tid	Sex	Age	Annual_Income	Category	Property
E101	M	35	10 Lakhs	Govt. Employed	15 Lakhs
E102	M	40	15 Lakhs	Self_Employed	25 Lakhs
E103	F	36	4 Lakhs	Self_Employed	10 Lakhs
E104	M	32	15 Lakhs	Govt. Employed	5 Lakhs
E105	F	25	5 Lakhs	Self_Employed	----

Table-1 Sample Dataset

Concept hierarchies for nominal attributes are constructed automatically based on discretization technology[6]. In order to satisfy multiple user's needs. More then one concept hierarchy can be defined. From the sample dataset given in table1, the concept hierarchy is generated as show in figure-2.

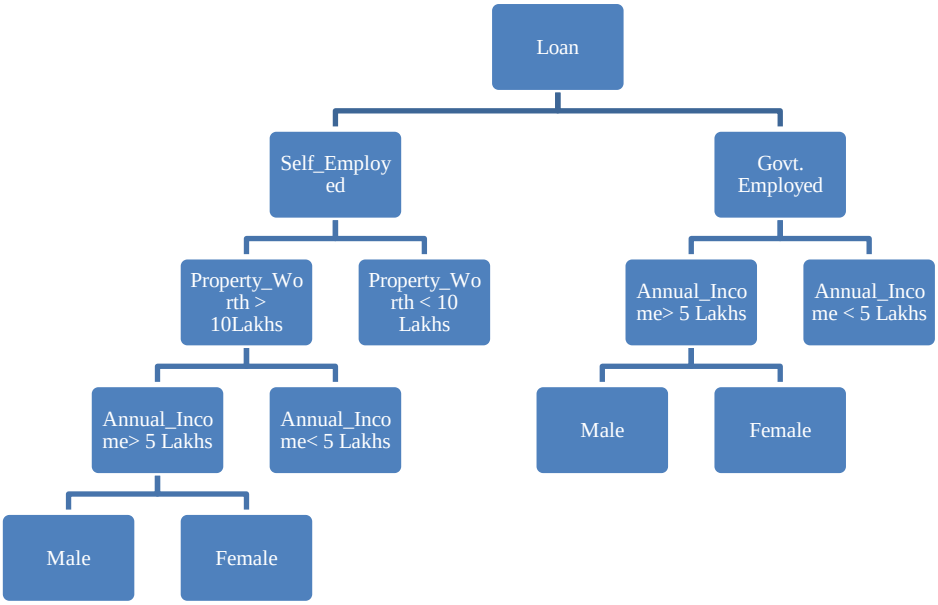

Figure-2 Concept Hierarchy for Sample Dataset

Combined value is integrated and identified each record based on the integer value which identifies an interesting rule as well as reduces the time complexity. Database with m-records and n-attributes requires m x 2^n computational steps for enumerating the attribute subset. Finally interesting rules are identified based on the height value.

Algorithm

Input: Dataset composed of N-Tuples, minsup, minconfidence and number of attributes.

Output: multidimensional association rules R

Step 1: Select particular set of attributes for the specified database.

Step 2: Let C_t be a set of constraints defined on these attributes

Step 3: Generate concept hierarchy for attributes

Step 4: Compute data conversion based on quantitative and binary values

Step 5: Associates Multi-Dimensional Relationship and find the relationship between multiple relations.

Step 6: Generate the rules from data extraction

Step 7: Filter the rules which satisfy the threshold limit attribute weight [(MinConfidence + Lift)/2].

IV. CONCLUSION

Mining frequent itemset and filtering an efficient itemsets for the association rule mining from large transactional database is a critical task. Many approaches have been proposed to extract the frequent itemsets for maximum scope. The main goal of this research considers the efficient rule generation and time complexity. This paper suggested as approach for reducing the time complexity as well as generates an efficient rule generation based on the threshold limit Weight. In future, this algorithm will be implemented and compresses the performance with existing algorithm.

References

1. R. Agrawal, T. Imielinski, and A. Swami. Mining association rules between sets of items in large databases. In Proc.1993 ACM-SIGMOD Int. Conf. Management of Data, Washington, D.C., May 1993, pp 207–216.
2. R. Agrawal and R. Srikant. Fast algorithms for mining association rules. In VLDBY94, pp. 487-499.
3. C.Borgelt. "Efficient Implementations of Apriori and Eclat". In Proc. 1st IEEE ICDM Workshop on Frequent Item Set Mining Implementations, CEUR Workshop Proceedings 90, Aachen, Germany 200.
4. J.S .Park, M.S.Chen and P.S.Yu. An effective hash based algorithm for mining association rules. In SIGMOD1995, pp 175-186.
5. J. Han, J. Pei, and Y. Yin. Mining Frequent Patterns without Candidate Generation (PDF), (Slides), Proc. 2000 ACM-SIGMOD Int. May 2000.
6. E.Ramaraj and R.Sridevi A general Survey on multidimensional and Quantitative Association Rule mining Algorithms. International Journal of Engineering Research and Applications(IJERA) 2013.
7. A.B.M.Rezbaul Islam, Tae-Sun Chung An Improved Frequent Pattern Tree Based Association Rule Mining Techniques Department of Computer Engineering Ajou University Suwon, Republic of Korea.
8. E. Ramaraj and N. Venkatesan, ― Bit Stream Mask Search Algorithm in Frequent Itemset Mining,‖ European Journal of Scientific Reasearch,‖ Vol. 27 No.2 (2009),
9. G. Grahne, J. Zhu, Fast algorithms for frequent itemset mining using FP-Trees, IEEE Transactions on Knowledge and Data Engineering 17 (10) (2005) 1347–1362.
10. J. Han, H. Cheng, D. Xin, X. Yan, Frequent pattern mining: current status and future directions, Data Mining and Knowledge Discovery (2007). 10th Anniversary Issue.
11. J. Han, J. Pei, Y. Yin, Mining frequent patterns without candidate generation, in: Proceedings of the 2000 ACM SIGMOD International Conference on Management of Data, 2000, pp. 1–12.
12. T.-P. Hong, C.-W. Lin, Y.-L. Wu, Incrementally fast updated frequent pattern trees, Expert Systems with Applications 34 (4) (2008) 2424–2435.
13. H. Huang, X. Wu, R. Relue, Association analysis with one scan of databases, in: Proceedings of the IEEE International Conference on Data Mining, 2002,pp. 629–632.
14. G. Grahne, and J. Zhu, "Efficiently using prefix-trees in mining frequent itemsets", IEEE ICDM Workshop on Frequent Itemset Mining Implementations, 2003.
15. C. Lucchese, S. Orlando, and R. Perego, "DCI_CLOSED: a Fast and Memory Efficient Algorithm to Mine Frequent Closed Itemsets", IEEE Transaction On Knowledge And Data Engineering, Vol. 18, No. 1, PP. 21-35, 2006.
16. D.Lin, Z. M. Kedem, Pincer-Search: A New Algorithm for Discovering the Maximum Frequent Itemset, EDBT Conference Proceedings, 1998, pages 105-110.

AUTHOR(S) PROFILE

Dr. K. Kavitha, is working as an Assistant Professor in Department of Computer Science at Mother Teresa Women's University, Kodaikanal. She has over 13 years of teaching experience. She has completed her MCA and Ph.D from Madurai Kamaraj University, Madurai in 2002 and 2014 respectively. She obtained her M.Phil in Computer Science from Bharathidasan University, Trichy in 2008. She has published around 15 Research Papers in peer received journals and attended National/ International Conferences. She is currently serving as a Member of International Association of Engineering. Her Research Area includes Data Mining, Cloud Computing and Mobile Computing.

