ISSN (e): 2250 – 3005 || Volume, 07 || Issue, 12 || December – 2017 ||
International Journal of Computational Engineering Research (IJCER)

A Survey on Optimization Techniques In Association Rule Mining Algorithms For Event Scheduling

[1]Dr.J.Chockalingam M.Sc.,M.B.A.,M.Phil.,Ph.D.,
[2]C.Seetharaman M.Sc.,M.Phil.,B.Ed.,(Ph.D)
[1]*Associate Professor Research Supervisor, Computer Science department Khadir Mohideen College, Adirampattinam-India*
[2]*Associate Professor Govt.Arts College(A), Kumbakonam Research Scholar, Computer Science department Khadir Mohideen College, Adirampattinam-India*
Corresponding Author: [1]Dr.J.Chockalingam M.Sc.,M.B.A.,M.Phil.,Ph.D.,

ABSTRACT
Analysis of large databases for market basket, frequent pattern identification, heuristics detection and event scheduling based on historical data is a difficult task in current day to day emerging situations. The transactional databases and relational databases are needed to be analyzed to accomplish the pattern identification. Association rule mining focuses on the analysis of historical data to perform the scheduling. The support and confidence values obtained from the ARM are used to perform the validation of ARM through the various techniques. Here the Association Rule Mining algorithms are discussed to make the worthiness of the predefined algorithms. The many techniques were proposed and validated for ARM using different kinds of datasets yet so far. Thus the predefined techniques are performed well however the optimization of analysis is required to speed up and improve the effectiveness of the algorithms and its techniques. In this paper, the optimization techniques like Particle Swarm Optimization, Ant Colony Optimization and Artificial Bee Colony with its features are discussed through the existing works.
Keyterms: Association Rule Mining, Frequent Pattern Identification, Heuristics Detection, Optimization Techniques

I. INTRODUCTION

In today's scenario, improving the analysis is an important task in the web mining and event scheduling. The data mining technologies has emerged due to the importance of analyzing the historical event and transactional data. It is very interesting to discuss about the Association Rule Mining and its features. Association Rule Mining[3] [4] is used to analysis the frequent if/then patterns and identify through the heuristics rule. The heuristics rule for the ARM is obtained through the two parts. One is antecedent (if) and consequent (then). The antecedent (if) of the rule is an item found in the training data and consequent (then) is an item that is found in the combination with the antecedent.

The Anomaly detection is the search for the events or itemsets obtained over the patterns based on the heuristics rule. These identified patterns are called anomalies and convert to actionable information in various application domains called outliers. With the help of Association rule mining [1] [3] [4], the events can be scheduled based on the patterns identified to predict the negative and positive situation for the education scenario. This type of analysis can be done in the web market data and transactional databases called Market Basket analysis. There are lot of routines can be used for the analysis based on Association rule mining. Thus the routines can be defined as predefined and effective algorithms such as AIS Algorithm, SETM Algorithm, Apriori Algorithm and Frequent Pattern growth (FP Growth Algorithm). Thus the algorithms are effectively composed and evaluated by the existing works. These algorithms are performed the analysis as much as well. The familiar algorithms which are used in above said algorithms are Apriori Algorithm and FP growth algorithm. For example, the heuristics rule for the Association Rule Mining as follows with an example for Market Basket Analysis.

Rule 1: If customer buys computer, then he/she will buy pen drive.
Rule 2: If customer buys milk, then he/she will buy bread/egg.

The effectiveness of these algorithms can be obtained through the analysis of numerical experiments based on the various datasets for some objective problem such as support, confidence, comprehensibility, interestingness and coverage. However the process of the analysis needs more effort when the historical or transactional database may take as large. Thus the algorithms are optimized through the optimization techniques [3] [5] [6]

[7] such as Particle Swarm Optimization, Ant Colony Optimization and Artificial Bee Colony Optimization. The further sections are incorporates the explanation of algorithms, overview of optimization techniques and the review about these techniques through the existing works.

II. ARM ALGORITHMS

As per discussions made in the proceeding section, the algorithms for the ARM [1] [3] [5] [4] have different strategies to obtain the event or pattern data. Association rules are if/then statements that help to find the relationships between the data and unrelated data in the database to obtain the patterns. Association rules are used to find the relationships between the test data that are frequently used in the database for transactions and frequently accessed based on the events.

In this section, the most important objective measures are used to find the efficiency of the processed algorithm. The support can be measured through the ratio of transactions in T containing the transaction records as X as below.

$$X\ (or)\ \sup(X) = \frac{|X(t)|}{|T|}$$

The support of the rule If X then Y can be computed and obtained through the following equation. The support of the rule X→Y is

$$Support(X \cup Y) = \frac{|X \cup Y|}{|T|}$$

The confidence measure should be calculated in next to the support value of the two data. The above equation calculates the support for the rule X→Y then the quality of the rule can be obtained through the ratio between the support value of number of transactions in the whole data. The confidence measure as follows.

$$Confidence = \frac{Support(X \cup Y)}{Support(X)}$$

In further section, the algorithms are reviewed with these measures are used with the obtained results from the same algorithm.

A. AIS Algorithm

The AIS Algorithm [4] was the unfamiliar but first proposed algorithm by the authors Agarwal, Imielinski and Swami for obtaining the Association rule Mining. It concentrates on the effectiveness of the quality of database analysis with the routines to process the decision support queries.In this AIS algorithm, the consequent of the rules only contain one item may generated like the rules as follows; for example, rules like X ∩ Y ⇒ Z can be generated but not the rules like X ⇒ Y ∩ Z. Thus the database should scan several times to obtain the frequent item sets through AIS. To make this algorithm more effective, the validation or estimation method was introduced to prune those itemsets which avoid the unwanted search and analysis done through the algorithm.

In [1], Trupti.A and Santosh.V were discussed in this paper about the algorithm and its working principle with transactional databases from the market itemsets.

B. SETM Algorithm

In this algorithm, the generated candidate itemsets based on the on-the fly as the data in the database may scan, but the counts can be obtained at the end of the next module. Then the procedure can be done as like as AIS Algorithm. The candidate itemset are generated as in sequential structure based on the support count of candidate itemsets is at the end. The support values can be same as the entries in the database is the main disadvantages of this algorithm and also in AIS Algorithm.Like above this algorithm also explained in [1], the authors also convey the disadvantages of these algorithms and need of newly proposed algorithm as follows in the further sections.

C. Apriori Algorithm

The Apriori Algorithm is a most familiar and plays a vital and useful role in mining association rules discovered by the Agrawal in 1994. The Apriori Algorithm can be used in two different ways. One is identifying the frequent itemsets mining with candidate subset generation and another one is frequent itemsets mining without candidate subset generation. The Apriori generates the candidate subsets from the transactional database T, where each transaction T_{id} is set of items. Let I be the set of items $I=\{I_1,I_2,I_3,\ldots I_n\}$. An itemset contains k items in the k-itemset. If obtained k-itemset contains minimum support (Min_Sup) then it is a frequent k-itemset, denoted by I_k . This algorithm generated a set of candidates, which is candidate k-itemsets, by C_k. If the candidate itemses obtains the minimum support the it is called frequent itemsets.

Algorithm: Apriori Algorithm

Input: Items I, $C_k=\{\}$, Min_sup

Output: Candidate Generation C_k

Procedure:
1. L_1=find the frequent itemsets in T
2. For k=2 while $L_{k-1}\neq\Phi$ start
3. C_k=apriori_gen(L_{k-1});
4. For each transactions t Є T start
5. C_t=subset(C_k,t);
6. For each candidate c Є C_t start
7. c.count++; end;
8. end;
9. L_k={cЄC_k|c.count≥min_sup}
10. end
11. return L=$U_k L_k$

Function Apriori_Gen(L_{k-1}: Frequent(K-1)-Itemsets)
1. for each itemset l_1ЄL_{k-1} start
2. for each itemset l_2ЄL_{k-1} start
3. If($l_1[1]=l_2[1]$)^($l_1[2]=l_2[2]$)^…^
($l_1[k-1]<l_2[k-1]$) then
4. C=$l_1\times l_2$; //join step
5. If has infrequent itemsubset(c, l_{k-1}) then
6. delete c; // prune step
7. else add c to C_k;
8. end
9. return C_k;

This above algorithm [4] [1] performs the operations of the candidate generation and frequent itemset identification. In [4], Han and Kamber describe the algorithm with a numerical example and also explain the join and prune step to obtain the frequent itemsets.

D. Frequent Pattern Growth Algorithm

The FP-Growth [4] [1] is another way to find frequent item sets without generating the candidate subset. By this way, it seems to be improved because of using divide and conquer strategy. However the improved performance obtained through this algorithm, it does not hold the FP-tree for the large databases in main memory. It is a very simple task in the algorithm. First it takes the transaction database as input and compresses to create an FP-Tree instance to represent frequent items. Next to this, based Divide and Conquer Method, it divides the compressed data into set of conditional databases, each item associated with one frequent pattern. At last, mining takes place in each such database separately. Using this technique, here the search costs will be reduced and looking for easiest patterns recursively.

It needs to construct the FP-Tree based on the quantitative information about frequent patterns in a database [1].

Algorithm: FP-Tree Construction
Input: Transaction Database D
Output: Frequent pattern Tree for D
Procedure:
1. Scan the transaction database DB
2. Collect the set of frequent items and discard the infrequent data items.
3. Sort the collected frequent items in descending order based on the support value.

III. OPTIMIZATION TECHNIQUES

The optimization techniques are use to expedite the process of rule mining and analysis. Here the familiar and most used optimization techniques [3] [1] [9] are (PCO) Particle Swarm Optimization, (ACO) Ant Colony Optimization and (ABC) Artificial Bee Colony Optimization.

IV. REVIEW ABOUT OPTIMIZATION TECHNIQUES

In this section, the optimization techniques and its features are discussed based on the existing works. Trupti et.al. [1] and Parmjeet Kaur [3] et.al discussed and reviewed the optimization techniques in the association rule mining. From their observations, the routines for the techniques are analyzed. In [11], Djenouri et al, proposed a novel hybrid algorithm called HBSO-TS for ARM based on two familiar metaheuristis algorithms Tabu Search and Bees Swarm Optimization. The Bee Swarm Search will analyze the search space in such away to cover

most of its space covered and the local exploration of each bee is computed by tabu search. Djenouri et al, [12] again presents an enhanced algorithm based on Genetic Algorithm and Bees Swarm Optimization at 2012. In need of handling and analyzing the large amount of data as some of existing and his existing algorithm in reasonable time. From the results obtained over this paper, BSO-ARM achieved slightly better than all the genetic approaches. Apart from that, BSO-ARM is more time consuming rather than the others. The two important operations like determination search area and neighbourhood search obtained by BSO Algorithm permit to improve the quality based on the performance but it requires a considerable computation time and is considered as future work.

S. Olafsson et al, [13] (2006) reviews and reveals the operation and research in data mining. The existing contributions of optimization methods in data mining touch on almost every part of the data mining process, from data visualization and pre-processing, to inductive learning, and picking up the best model after learning. These methods still require significantly more research. In [14], Binitha S et al, reveals the different Bio-Inspired Algorithm used in ARM optimization. There still remain significantly challenging tasks for the research community to address for the realization of many existing and most of the emerging areas in technology. M.A. Nada [15] hybridized the Ant Colony algorithm and Genetic programming algorithm to incorporate the movements of ant towards solution. The genetic programming paradigm allows the hyper evolution of programs in computers which can perform slightly changed alternative computations accustomed on the upshot of transitional calculations.

V. CONCLUSION

Thus the paper concludes the review about the optimization techniques over the Association Rule Mining. In this survey we have studied mostly used technique that is Apriori Algorithm and FP-Growth Algorithm for finding the association between frequently occurred itemsets from large set of transactions. The need of the optimization techniques are also discussed and surveyed based on the existing works.

REFERENCES

[1]. Trupti A.kumbhare and Santosh.V.Shobe, "An Overview of Association Rule Mining Algorithms", Int. Journal of Computer Science and Information Technologies, Vol.5, No.1, pp.927-930, 2014.

[2]. Imam Tahyudin and Hidetaka Nambo, "The Combination of Evolutionary Algorithm Method for Numerical Association Rule Mining Optimization", Int. Conf. on Advances in Intelligent Systems and Computing, Vol. 502.

[3]. Parmjeet Kaur, Usvir Kaur and Dheerendra Singh, "Optimization Techniques in Association Rule Mining: A Review", International Journal of Current Engineering and Technology, Vol.4, No.3, June 2014.

[4]. Han. J, Kamber.J, and Pei.M, "Data Mining: Concepts and Techniques", Elsevier 2nd Edition.

[5]. Badri Prasad Patel, Nitesh Gupta, Rajneesh K and Y.K.Rana, "Optimization of Association Rules Mining Apriori Algorithm Based on ACO", International Journal on Emerging Technologies, Vol.2, No.1, pp.87-92, 2011.

[6]. Neeraj Rabeja and Ravish Kumar, "Optimization of Association Rule Learning in Distributed Databse using Clustering Technique", Int. Journal of Scientific and Research Publications, Vol. 2, No.12, December 2012.

[7]. Swati Shrivastava and Vikram Rajput, "Particle Swarm Optimization Based Association Rule Mining", Int. Journal of Computer Applications, Vol.5, No.5, August 2015.

[8]. G Dhanore and S.K.Chaturvedi, "An Optimization of Association Rule Mining for large database using K-Map and Genetic Algorithm: A Review", Int. Conf. on Electronics and Communication System, 2014.

[9]. Manish Saggar, Ashish K Agarwal and Abhimanyu Lad, "Optimization of Association Rule Mining using Improved Genetic Algorithms", IEEE Int. Conf. on Systems, Man and Cybernatics, 2004.

[10]. Liu, B., Hsu, W. and Ma, Y. (August 15-18, 1999), Mining association rules with multiple minimum supports, ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (KDD-99), San Diego, CA, USA.

[11]. Djenouri Y, Drias H., Habbas, Z., &Mosteghanemi H.(Dec. 2012), Bees Swarm Optimization for Web Association Rule Mining. Web Intelligence and Intelligent Agent Technology (WI-IAT), 2012 IEEE/WIC/ACM International Conferences, 3, pp-142 – 146.

[12]. Djenouri Y., Amine Chemchem (Aug. 2013), "A Hybrid Bees Swarm Optimization and Tabu Search Algorithm for Association", Nature and Biologically Inspired Computing (NaBIC), 2013 World Congress, pp-120 -125.

[13]. S. Olafsson, X. Li, S. Wu, "Operations research and data mining", Elsevier, 2016.

[14]. Binitha S, S Siva Sathya, "A Survey of Bio inspired Optimization Algorithms", International Journal of Soft Computing and Engineering (IJSCE), 2(2), ISSN: 2231-2307, May 2012.

[15]. Nada M. A. Al Salami, "Ant Colony Optimization Algorithm",UbiCC Journal, 4(3), pp 823-826, 2009.

International Journal of Computational Engineering Research (IJCER) is UGC approved Journal with Sl. No. 4627, Journal no. 47631.

Dr.J.Chockalingam M.Sc.,M.B.A.,M.Phil.,Ph.D. "A Survey on Optimization Techniques In Association Rule Mining Algorithms For Event Scheduling." International Journal of Computational Engineering Research (IJCER), vol. 07, no. 12, 2017, pp. 43-46.