A Study on Milestones of Association Rule Mining Algorithms in Large Databases

Saravanan Suba
Assistant Professor of Computer Science
Government Arts College
Paramakudi,TN,India

Chistopher.T, PhD.
Assistant Professor of Computer Science&Head
Government Arts College
Udumalpet,TN,India

ABSTRACT
Data mining helps in doing automated extraction and generating predictive information from large amount of data. The association rule mining is one of the important area of research in Data mining. The Association rule mining identifies the useful associations or relationship among big set of data items. In this paper, we provide the important concepts of Association rule mining and existing algorithms and their effectiveness and drawbacks. The references provided in this paper covered the main theoretical issues and guiding the researcher in an interesting research directions that have yet to be discovered.

Keywords
Data Mining, Association Rule Mining, Apriori, FP-growth, Frequent Item sets

1. INTRODUCTION
Nowadays there has been an exponential growth in the generation and manipulation of electronic information as more and more operations are computerized. Any organization or enterprise have started to realize that the information accumulated over years is an important strategic asset and they also realize that there is potential business intelligence hidden in the large amount of data. For that, what these enterprises want is a technique that permits them to extract the most valuable information from accumulated data. The field of data mining offers such techniques which evaluating the current data and inferring hidden information that would be useful in future prediction, pattern recognition and decision making [1].

Data mining is a collection of techniques for effective automated discovery of formerly unknown, valid, novel, useful and understandable pattern in large databases [1]. The pattern must be usable so that they can be used in the enterprise's decision making process. Data mining is also seen as an important step in the overall process of knowledge discovery which composed of various segments. Data Cleaning frees noise and other inconsistent data which are present in the input database. Since the input data base could be composed of data from multiple sources, Data Integration is employed to merge data from different sources. This is

otherwise called Data Warehouse. Data Selection phase finds the specific data mining task relevant data in the input data base. Data Transformation phase which transfer the input data into format suitable for data mining. The specific data mining task that employs clever methods or algorithm for mining is carried out. Next, the interesting pattern will be selected from set of pattern mined from previous step. The last step of KDD (Knowledge Discovery In Data Base) process is the Presentation of the Discovered Knowledge in the user friendly format [2].

The Data Mining techniques or tasks can be generally classified as descriptive or predictive. Descriptive mining refers to the method in which the essential characteristics or general properties of the data in the data base are depicted. The descriptive techniques involve task like Clustering, Association and Sequential Mining[3]. Predictive data mining tasks are those that perform inference on input data to arrive at hidden knowledge and make interesting and useful prediction[2]. The predictive mining techniques involve tasks like Classification, Regression and Deviation[3].

Data Mining is motivated by decision support problems faced by most business organizations and is described as essential area of research [4]. Key research issues or challenges in Data Mining are performance, mining methodology, user interaction and data diversity . So the data mining algorithm and methodologies must be competent and scalable well to the size of data base and their execution times[2].

One of the most popular descriptive Data Mining techniques is association rule mining [3]. Since its introduction[5], Association Rule Mining has become one of the core Data Mining tasks and has attracted tremendous interest among Data Mining researches and practitioners[6].

Association Rule Mining could be decomposed into two sub problems, mining large item set (i.e. frequent item sets) and the generation of association rules[2]. Two statistical measures that control the process of association rule mining are support and confidence. For an association rule $X \rightarrow Y$ and number of transactions is denoted as N, the support and confidence can be mathematically represented as follows

$Support(X \rightarrow Y) = \sum(XUY)/N$ and
$Confidence(X \rightarrow Y) = \sum(XUY)/ \sum X$.

The entire process of association and pattern mining is controlled by user specific parameter, namely minimum support and confidence[2]. The Association Rule Mining(ARM) task was first presented by Agrawal et al. [5] to discover interesting relationships among items in market basket transactions. Since its inception, extensive studies have been being conducted to address various conceptual, implementation and application issues pertaining to the association analysis task.

Research in conceptual issues in Association Rule Mining is focused primarily on developing a framework to define the theoretical underpinnings of association analysis and extending the formulation to handle new type pattern. Research in implementing issues in ARM involve mixing the mining capability into existing data base technology, developing competent and scalable mining algorithm, handling user specified or domain specific constraints and post processing the extracted pattern. Research in application issues in Association Rule Mining includes marketing,

medicine, electronic commerce, classification, clustering, web mining, bio informatics, finance, student database and many more[1].

It is clear that ARM is an interesting problem with numerous applications and it is undirected or unsupervised data mining over variable length data and it produce clean and understandable results.The research in association rule mining algorithm can be made effective than Apriori algorithm. The effectiveness of association mining rule in pattern discovery can be derived based on computational model and algorithm used. The computation model can be serial or parallel and online or batch. This research mostly focuses on implementing issues of association rule mining algorithms. Existing algorithms are reviewed and challenges pertaining to this domain will be raised.

This paper is organized as follows: Section 2 presents the background and related work of association rule mining, Section 3 compares and discusses the various existing association rule mining algorithms according to its effectiveness and improvements, Section 4 explains the challenging problems in association rule mining and proposed method and section 5 describes the conclusion.

2. MILESTONES OF ASSOCIATION RULE MINING ALGORITHMS

In this section , the association rule mining problem is presented in detail. Several issues in association rule mining have been elaborated together with classic algorithms. ARM aims to extract interesting correlations, frequent patterns, associations or casual structures among sets of items in the transaction databases or other data sources.

The research activities in this area revolve around incorporating the mining capability into existing database technology, Developing competent and scalable algorithms, handling user specific or domain specific constrains and Post processing of extracted patterns.

2.1 AIS Algorithm

The AIS (Agrawal, Imielinski and Swami) algorithm was the first algorithm proposed to generate association rule by Agrawal et al. in 1993 [5]. It focus on improving the quality of databases together with necessary functionality to process decision support queries.

The main disadvantage of the AIS algorithm is too many candidate itemsets that finally turned out to be small are generated, which needs more space and times. At the same time, this algorithm requires too many scans over the whole database to generate large itemsets[7].

2.2 Apriori Algorithm

The exploitation of monotonic property of the support of itemsets and confidence of association rules had been created a situation to enhance AIS algorithm and it was renamed as Apriori[8]. Apriori is a best improvement in the history of association rule mining. Apriori algorithm was first proposed by Agrawal and Srikant[9]. The AIS is just a straightforward method that requires many passes over the database and generating many candidate itemsets and storing counters of each candidate while most of them turn out to be not frequent. Apriori works efficiently during the candidate generation process for two reasons, Apriori employs a different candidates generation method and a new pruning technique.

In the process of finding frequent itemsets, Apriori avoids the effort of wastage of counting the candidate itemsets that are

known to be infrequent. The candidates are produced by joining among the frequent itemsets level-wisely and are filtered according the Apriori property. As a result, the number of remaining candidate itemsets ready for further support checking becomes much smaller, which enormously reduces the computation, I/O cost and memory requirement. Detail of the Apriori-gen and GenerateRules functions were expanded by Agrawal and Srikant in 1994[8]. Apriori algorithm still inherits the disadvantage of scanning the whole data bases many number of times. Based on Apriori algorithm, many new algorithms were designed with some alterations or improvements. Generally there were two methods: one is to reduce the number of passes over the whole database or replacing the whole database with only part of it based on the current frequent itemsets and another approach is to explore different kinds of filtering techniques to make the number of candidate itemsets much smaller.

2.3 AprioriTid and Apriori Hybrid

Apriori suffers from limitations of vast number of repeated input scans. Since it is level wise algorithm, it needs separate scans of the input database and over the entire frequent itemset mining process, this become tedious and is severe limitation.

One of variations is algorithm AprioriTid (transaction id), emphasizing the fact that transctions in the database are replaced by candidate itemsets that occur in that transaction. This variation of Apriori performs well at higher level where as the conventional Apriori performs better at lower levels[2].

To avoid the disadvantages of AprioriTid algorithm, the concept of Apriori-hybrid which combined the advantageous features of Apriori and AprioriTid was proposed by Agrawal and Srikant in 1994[9].

SETM (SET-oriented Mining of association rules) [10] was constantly outperformed by AIS. AprioriTid performed equivalently well as Apriori for lesser problem sizes. However, performance reduced twice slow when applied to big problems.

2.4 Hash Based Itemset Counting

The DHP(Direct Hashing and Pruning) algorithm is an effective hash-based algorithm for the candidate set generation and it was suggested by park et al. in 1995[11]. Note that the DHP algorithm has two major features: One is its efficiency in generation of large itemset and other is effectiveness in lessening on transaction data base size. Hence, A hash technique is very efficient in generating the candidate itemsets, in particular for the large two itemsets, thus really improving the performance bottleneck of the entire process.

An effective Direct Hashing and Pruning [DHP] algorithm was proposed by Soo et al. in 1997[12] for mining the association rules. This algorithm employs effective pruning techniques to progressively reduce the transaction database size. DHP uses a hashing technique to screen the ineffective candidate frequent 2 itemsets. DHP also avoids database scans in some passes as to reduce the disk I/O cost involved.

En et al. [13] introduced another novel hash-based approach for mining frequent itemsets over data streams. The algorithm compresses the information of all itemsets into a structure with a stable hash-based technique. This approach expertly summarizes the information of the whole data stream by using a hash table to estimate the support counts of the non-frequent

itemsets and keeps only the frequent itemsets for speeding up the mining process. Another algorithm Inverted Hashing and Pruning (IHP) proposed by John et al. in 2000[14]. It was developed for mining association rules between words in text databases. It has been demonstrated that the IHP algorithm has better performance for large text databases.

2.5 Dynamic ItemSet Counting Algorithm

This algorithm suggested by Brin et al. in 1997 [15] aimed at reducing the number of database scan by dividing the database into intervals of specific sizes. In effect, the algorithm reduces the number of database scans to 1.5 as contrasting to 3 scans required by level-wise Apriori approach. The algorithm is based on the overall principle of counting for itemsets when ever it is optimal rather than having to wait for completion of the previous pass.

2.6 CARMA Algorithm

The performance of the DIC algorithm is deeply dependent on the real delivery of data in the database. The CARMA(Continuous Association Rule Mining Algorithm) proposed by Hidber et al. in 1999[16]. CARMA results in more candidate sets compared to Apriori and DIC despite offering the flexibility of allowing the user to alter minimum support threshold values during the algorithm's execution.

2.7 Sampling Approach

Sampling is a dominant data reduction technique that has been applied to a variety of data mining algorithms for reducing computational overhead. In the context of association rules, sampling can be used to collect quick preliminary rules. This may help the user to direct the data mining process by purifying the criterion for interesting rules. Sampling can speed up the mining process by more than an order of magnitude by reducing I/O costs and hugely shrinking the number of transaction to be considered. The validity of the sample is determined by two characteristics the size of the sample and the quality of the sample. The quality, in the context of statistical sampling techniques, refers to whether the sample captures the characteristics of the database [17].

Toivonen [18] proposed an association rule mining algorithm using sampling in 1995. This approach uses two passes. During pass1 a sample of the database is obtained and all associations in the sample are found. These results are then validated against the entire database. To maximize the effectiveness of the whole approach, the author makes use of lowered minimum support on the sample. Since the method is probabilistic (i.e. dependent on the sample containing all the relevant associations) not all the rules may be identified in this first pass. Those associations that were deemed not frequent in the sample but were actually frequent in the entire dataset are used to construct the complete set of associations in pass2.

Parthasarathy proposed an efficient method to progressively sample for association Rules[19] in 2002. His method relies on a novel measure of model accuracy (self similarity of associations across progressive samples), the identification of a representative class of frequent itemsets that mimic (extremely accurately) the self similarity values across the whole set of associations and an efficient sampling methodology that hides the overhead of attaining progressive samples by overlapping it with valuable computation.

Chuang et al. [20] presented another progressive sampling algorithm, called Sampling Error Estimation (SEE), which aims to identify the suitable sample size for mining association rules. SEE has two benefits. First, SEE is highly

efficient because an appropriate sample size can be determined without the need of executing association rules. Second, the identified sample size of SEE is very accurate, meaning that association rules can be highly efficiently executed on a sample of this size to attain a sufficiently accurate result. Especially, if data comes as a stream flowing at a faster rate than can be processed, sampling appears to be the only choice. How to sample the data and how big the sample size should be for a given error bound and confidence level are key issues for specific data mining tasks. Li and Gopalan [21] derive the sufficient sample size based on central limit theorem for sampling large datasets with replacement in 2004. A number of studies were conducted to propose effective methods for mining association rules by reducing either the CPU computation time or the disk access overhead. Some studies considered the usage of sampling methods for reducing the processing overhead [22,23] .

2.8 Partitioning Approach

The partitioning algorithm proposed by savasere et al. [24] in 1995 adopts a divide and conquer method to process of frequent itemset mining. The input database is separated in to several disjoined parts (partitions) and frequent items with in the partitions are generated. Later, the frequent itemsets obtained from various partitions are combined and resulting in a superset of all frequent itemset in the complete database. Actual counts of the frequent itemset are calculated during a second scan of the database. So the partitioning may increase the performance of finding large itemsets in numerous ways. By using partitioning, parallel and/or distributed algorithms can be easily created where each partition could be handled by a separate machine.

2.9 Parallel Association Rule Mining

Cheung et al. [25] proposed an algorithm called FDM in 1996. FDM is a parallelization of Apriori to shared machines, each with its own partition of the database. At each level and on each machine, the database scan is accomplished independently on the local partition. Distributed Pruning is then done.FPM (Fast Parallel Mining) for Association rule mining has been proposed [26] in 1998. It adopts Count Distribution method and has incorporated two powerful candidate trimming techniques. It has a simple communication scheme which performs only one round of message exchange in each iteration.

Parthasarathy et al. [27] have presented an excellent survey on parallel association rule mining with shared memory architecture covering various challenges and approaches adopted for parallel data mining in 2001.Tang and Turkia [28] suggested a parallelization scheme which can be used to parallelize the effective and fast frequent itemset mining algorithms based on FP-trees.

2.10 FP Growth Approach

The previously debated methods have primarily adopted a breadth first strategy towards the issue of frequent itemset mining. The size of the candidate itemsets can be reduced by adopting a depth first logic as done by FP growth algorithm[2]. In several cases, the Apriori algorithm significantly lessens the size of candidate sets using the Apriori property. However, it has the following weaknesses: generating a big number of candidate sets, repeatedly scanning the data base and checking the candidates by pattern matching.

Han et al. [29] developed an FP growth algorithm that mines the full set of frequent itemsets without candidate generation.

FP-growth algorithm is based on the divide and conquer principle. The main disadvantage in FP-tree is that the construction of the frequent pattern tree is a time consuming activity. More over, FP-tree based approaches do not offer flexibility and reusability of computation during mining process.Performance studies show that the method substantially reduces search time. Other notable algorithms include Tree Projection [30] in 2001, H-mine[31] which uses hyper structure in 2001, Pattern growth mining using top down and bottom up traversals by Liu et al. [32] in 2002 and an array based implementation of prefix tree structure for efficient pattern growth mining by Grahne and Zhu[33] in 2003.

2.11 Association Rule Discovery With Constrains

Several data mining methods consist in discovering patterns frequently occurring in the source dataset. Typically, the aim is to discover all the patterns whose frequency in the dataset exceeds a user-specified threshold. However, very often users want to limit the set of patterns to be discovered by adding extra restrictions on the structure of patterns. Data mining systems should be able to exploit such restrictions to speedup the mining process. Techniques applicable to constraint-driven pattern discovery can be classified into the following groups: post-processing (filtering out patterns that do not fulfill user-specified pattern constraints after the actual discovery process), pattern filtering (integration of pattern constraints into the actual mining process in order to produce only patterns satisfying the constraints) and dataset filtering (restricting the source dataset to objects that can possibly have patterns that satisfy pattern constraints).

Srikant et al. in 1997 and Ng et al. in 1998 suggested some algorithms with faster association rule mining by incorporating item constraints on the process of generating frequent itemsets[34].

Wojciechowski and Zakrzewicz [35] focus on improving the effectiveness of constraint- based frequent pattern mining by using dataset filtering techniques. Dataset filtering conceptually transforms a given data mining task into an equivalent one operating on a lesser dataset. Tien Dung Do et al. [36] proposed a specific type of constraints called category-based as well as the associated algorithm for constrained rule mining based on Apriori. This method reduces computational complexity of mining process by passing most of the subsets of final itemsets.

2.12 Multi Level Association Mining

In actual life, for various applications, it is tough to find strong association rules between data items at low or primitive level of abstraction in multidimensional space [37]. While strong association rules produced at a higher concept level may be common sense to some users but it also can be novel to other users. Multiple level association rule mining is trying to mine strong association rules among intra and inter different levels of abstraction. Researches have been done in mining association rule at multiple concept levels [38,39,40].

Mining association rules at basic level, in many cases, loose detailed information. Further it can show only general rules without ability of getting inside the rule. Data mining should also be available for mining association rules at the multiple levels of abstraction. In association rules each transaction can be encoded based on dimension and levels.

In multiple level association rule mining, the items in an itemset are characterized by using a concept hierarchy. Mining happens at multiple levels in the hierarchy. At lowest levels, it might be that no rules may match the constraints. At highest levels, rules can be extremely general. Generally, a top-down method is used where the support threshold may be same or varies from level to level (support is reduced going from higher to lower levels) [41].

2.13 Negative Association Rule Mining

Typical association rules consider only items enumerated in transactions. Such rules are denoted to as positive association rules. Negative association rules also consider the same items, but in addition consider negated items (i.e. absent from transactions).

Negative association rules are valuable in market-basket analysis to identify products that conflict with each other or products that complement each other. Mining negative association rules is a hard task due to the fact that there are essential variances between positive and negative association rule mining. The researchers attack two key difficulties in negative association rule mining: (i) how to effectively search for exciting itemsets and (ii) how to efficiently identify negative association rules of interest. Brin et al. [42] stated for the first time in the literature the notion of negative relationships. Their model is chi-square based. They use the statistical test to confirm the independence between two variables. To determine the nature (positive or negative) of the relationship, a correlation metric was used.

In [43] the authors presened a new notion to mine strong negative rules. They combine positive frequent itemsets with domain knowledge in the form of taxonomy to mine negative associations. However, their algorithm is tough to generalize since it is domain dependant and needs a predefined taxonomy. A similar approach is defined in [44].

Wu et al. [45] derived a new algorithm for generating both positive and negative association rules. They add on top of the support-confidence framework another measure called mininterest for a better filtering of the frequent itemsets generated in 2002.

A innovative method has proposed [46] in 2004. In this, mining both positive and negative association rules of interest was decomposed into the following two sub problems, viz., (1) generate the set of frequent itemsets of interest (PL) and the set of infrequent itemsets of interest (NL) (2) Extract positive rules in PL and negative rules in NL. To generate PL and NL, they developed three functions namely, fipi(), iipis() and CPIR().

In [47] authors considered another frame-work called correlation analysis that adds to the support-confidence. In [47], they joined the two phases (mining frequent itemsets and generating strong association rules) and generated the significant rules while analyzing the correlations within each candidate itemset. This avoids evaluating item combinations redundantly. Indeed, for each generated candidate itemset, they computed all possible combinations of items to analyze their correlations. At the end, they retain only those rules generated from item combinations with strong correlation. If the correlation is positive, positive rules is discovered. If the correlation is negative, negative rules are discovered.

An innovative approach was proposed [48] to produce positive and negative association rules 2006. Authors generated negative rules without adding additional interesting measure(s) to support confidence frame work.

Recently B.Ramasubbareddy et al. [49] proposed an algorithm MPNAR to generate both positive and negative association rules in 2010. Their method uses the Yule's coefficient and works in the support-confidence framework to generate stronger positive and negative rules. Its performance was compared with MLA [47] on a synthetic dataset. By varying minimum support and minimum confidence and found that this algorithm performs better than MLA stated in [47].

2.14 Other Important Association Rule Mining Algorithms

Rapid association rule mining[50],Generalized association rule mining[51,52,53] ,Fuzzy association rule mining [54] Mining, Distributed association rule mining[55], Association rule mining using multi criteria decision methods (Global profit weight method)[56], Frequent item sets using vertical layout[57], Maximal and closed frequent pattern mining algorithms[57], Multi dimensional and Quantitative association rule mining algorithms [57], Sequential association rule mining algorithms [57], Incremental association rule mining algorithms [57], Image association rule mining[3] and Association rule mining for clustering[58,59,60] are seen in the literature.

3. COMPARATIVE ANALYSIS

Table 1 ARM Algorithms comparison

Algorithm Name	Search Type	Number of Passes/Scans	Data Structure
AIS[5]	Breadth First Search	K+1	List
Apriori[8]	Breadth First Search	K+1	Hash Tree + Hash Table
Apriori Tid[2]	Breadth First Search	K+1	Tree
Apriori Hybrid[9]	Breadth first search	K+1	Tree
SETM[10]	Depth First Search	K+1	Not specified
DHP[11]	Breadth First Search	Multiple	Hash table
DIC[15]	Breadth/ depth first search	Multiple	Trie
CARMA[16]	Breadth/ depth first search	2	Hash Table
Sampling[18]	Not specified	2	Not specified
Partitioning[24]	Breath first search	2	Hash table
FP-Growth[29]	Divide and conquer	2	FP-tree
Dynamic FP-Tree[61]	Divide and conquer	1	FP-tree

The comparison scheme provides a frame work which clearly shows the search type, number of scans required(k-represent number of items) and data structure of the various association rule mining algorithms. The above mentioned algorithms also compare using various metrics such as space and time needed to execute the algorithms. The performance of any algorithm evaluated by calculating time needed to execute the algorithm. The time requirements is estimated by counting the maximum number of scans required and maximum number of comparison operations. The above table clearly says that the Dynamic FP-Tree outperforms than remaining algorithms because it needs only one scan.

4. CHALLENGING ISSUES AND PROPOSED METHOD

Algorithms are available for retaining the association rules due to addition or deletion of transactions in the data base. Efficient algorithms are not available for mining incremental rules due to adding of more items.

There is a requirement for the improvement of effective distributed algorithms in order to speed up the computation activity and improve the overall performance.

Most of the algorithms available in the literature for mining frequent itemsets do not offer flexibility for reusing the computation during mining process. It still needs research to shrink the size of derived pattern and enrich the quality of retained patterns.

4.1 Proposed Method

The Common limitation of existing association rule mining algorithm is that all algorithms generate lot of rules for future predictions. Some of them are not actionable. Those rules should be filtered to get meaningful rules that are immediately actionable. In our proposed algorithm, we are trying to integrate user domain expert knowledge , ontology concept[62] and related interesting measures in the post processing step effectively than existing method in order to reduce the number of rules. So the found rules will be applied as easily as possible to improve any business.

5. CONCLUSION

It is concluded that association rule mining is an interesting pattern mining problem. The algorithms used are conceptually clear and resulting rules are pure and understandable. A decade from the important work of Agrawal et al. [5], association mining has become a field of research with different branches of specialization. The fundamentals of association mining are now well established with some exemptions. It highlights the fundamental principles within itemset identification and rule generation.

Association rule mining is being used as an significant tool for knowledge discovery and its applicability in other data mining tasks such as clustering and classification. This survey has delivered an organization of the significant fundamental contributions made within association mining research over this time and identify some gap available in the current knowledge for future research in this region.

6. REFERENCES

[1] G.K.Gupta, 2009 " Introduction to Data mining with Case Studies", PHI Learning private limited, New Delhi.

[2] N.P.Gopalan and B.Sivaselvan, 2009 " Data mining Techniques and Trends" PHI Learning private limited, New Delhi.

[3] S.Shankar and T.Purusothaman, 2009 "Utility Sentient Frequent Item set Mining and Association Rule Mining: A Literature survey and Comparative Study", International Journal of Soft Computing Applications ISSN: 1453-2277 Issue 4 , pp.81-95.

[4] Ashok Savasere , Edward Omieinski and Shankant Navathe, 1995 "An Efficient Algorithm for Mining Association Rules in Large Databases", Proceedings of the 21st International Conference on Very Large Data Bases, pp. 432 – 444.

[5] R.Agrawal, T.Imielinski, and A.Swami, 1993 "Mining Association Rules Between Sets Of Items In Large Databases", In proceedings of the ACM SIGMOD International Conference on Management of data, pp. 207-216.

[6] M. J. Zaki and C.J. Hsiao, October 1999 "CHARM: An efficient algorithm for closed association rule mining", Technical Report 99-10, Computer Science Dept., Rensselaer Polytechnic Institute.

[7] Sotiris Kotsiantis and Dimitris Kanellopoulos, 2006 "Association Rules Mining: A Recent Overview", GESTS International Transactions on Computer Science and Engineering, Vol.32, No: 1, pp. 71-82.

[8] Anurag Choubey, Ravindra Patel,J. L.Rana, May 2011 "A Survey Of Efficient Algorithms And New Approach For Fast Discovery Of Frequent Item Set For Association Rule Mining", International Journal of Soft Computing and Engineering.

[9] Rakesh Agrawal and Ramakrishnan Srikant, Sep'1994 "Fast Algorithms For Mining Association Rules In Large Databases", In Jorge B. Bocca, Matthias Jarke, and Carlo Zaniolo, editors, Proceedings of the 20th International Conference on Very Large Data Bases, VLDB, pp 487-499, Santiago, Chile.

[10] M. Houtsma, and Arun Swami, 1995. "Set-Oriented Mining for Association Rules in Relational Databases", IEEE International Conference on Data Engineering, pp. 25–33.

[11] Park, J. S, Chen, M.S and Yu P. S, 1995 "An Effective Hash Based Algorithm For Mining Association Rules" , In Proceedings of the 1995 ACM SIGMOD International Conference on Management of Data, M. J. Carey and D. A. Schneider, Eds. San Jose, California, pp.175-186.

[12] Soo J, Chen, M.S, and Yu P.S, 1997 "Using a Hash-Based Method with Transaction Trimming and Database Scan Reduction for Mining Association Rules" , IEEE Transactions On Knowledge and Data Engineering, Vol.No.5. pp. 813-825.

[13] En Tzu Wang and Arbee L.P. ChenData, " A Novel Hash-Based Approach For Mining Frequent Item-Sets Over Data Streams Requiring Less Memory Space", Data Mining and Knowledge Discovery, Volume 19, Number 1, pp 132-172.

[14] John D. Holt and Soon M. Chung," Mining of Association Rules in Text Databases Using Inverted Hashing and Pruning" Lecture Notes in Computer .

[15] S. Brin, R. Motwani, J.D. Ullman, and S. Tsur, 1997 "Dynamic Itemset Counting And Implication Rules For Market Basket Data", In Proceedings of the 1997 ACM SIGMOD, International Conference on Management of Data, volume 26(2) of SIGMOD Record, pp. 255–264. ACM Press.

[16] C. Hidber, 1999 "Online Association Rule Mining", In A. Delis, C. Faloutsos, and S.Ghandeharizadeh, editors, Proceedings of the 1999 ACM SIGMOD International Conference on Management of Data, volume 28(2) of SIGMOD Record, pp. 145–156, ACM Press.

[17] V.Umarani et al., 2010 "A Study on Effective Mining of Association Rules from Huge Databases", International journal of computer science and research,Vol .1,issue 1.

[18] Toivonen H, 1996 "Sampling large databases for association rules ", In VLDB Journal, pp. 134-145.

[19] Parthasarathy S, 2002 "Efficient Progressive Sampling for Association Rules", ICDM 2002, pp.354-361.

[20] Chuang K, Chen M and Yang W, Jun 2005 "Progressive Sampling for Association Rules Based on Sampling Error Estimation", Lecture Notes in Computer Science, Volume 3518, pp. 505 – 515.

[21] Li Y and Gopalan R, Jan 2004 "Effective Sampling for Mining Association Rules, Lecture Notes in Computer Science", Volume 3339, pp. 391 – 401.

[22] V.Umarani and M.Punithavalli, 2009 " Developing a Novel and Effective Approach for Association Rule Mining Using Progressive Sampling" In the proceedings of 2nd International Conference on Computer and Electrical Engineering (ICCEE 2009), vol.1, pp610-614.

[23] V.Umarani and M.Punithavalli , April 2010 "On Developing an Effectual Progressive Sampling Based Approach for Association Rule Discovery", In the proceedings of 2nd IEEE International Conference on Information and data Engineering (2nd IEEE ICIME 2010), Chengdu ,China .

[24] Savesere A, Omiecinski E, and Navathe S, 1995 "An Efficient Algorithm For Mining Association Rules In Large Databases", In Proceedings of 20th International Conference on VLDB.

[25] Cheung D, Han J, Ng V, Fu A and Fu, Y 1996, "A Fast Distributed Algorithm For Mining Association Rules", in Proceedings of 1996 International Conference on Parallel and Distributed Information Systems, Miami Beach, Florida, pp.31-44.

[26] Cheung. D and Xaio.Y, 1998 "Effect Of Data Skewness In Parallel Mining Of Association Rules", Lecture Notes in Computer Science, Volume 1394, pages 48-60.

[27] Parthasarathy.S, Zaki, M.J.J and Ogihara M, 2001 "Parallel Data Mining For Association Rules On Shared-Memory Systems", Knowledge and Information Systems: An International Journal, 3(1), pp.1-29.

[28] Tang P and Turkia M, 005 " Parallelizing frequent itemset mining with FP-trees. Technical Report titus.compsci.ualr.edu/~ptang/papers/par-fi.pdf", Department of Computer Science, University of Arkansas at Little Rock.

[29] Han,J And Pei.J, 2000 " Mining Frequent Patterns By Pattern Growth: Methodology And Implications", SIGKDD Explorations 2, 2, 14–20.

[30] Agarwal, R. Aggarwal, C and Prasad V, 2001 "A Tree Projection Algorithm For Generation Of Frequent

Itemsets",Iternational journal of Parallel and Distributed Computing.

[31] Pei.J, Han. J, and Lakshmanan L. V. S, 2001 "Mining Frequent Itemsets With Convertible Constraints", In Proceedings of the 17th International Conference on Data Engineering (ICDE'01) Heidelberg, Germany , IEEE Computer Society Press, pp. 433 – 442.

[32] Liu. J, Pan. Y , Wang. K And Han. J, 2002 "Mining Frequent Item Sets By Opportunistic Projection", In Proceedings of the Knowledge Discovery in Databases. Edmonton, Canada, Vol. 31. ACM Press, pp.97–102.

[33] Grahne G AND Zhu J, 2003 " Efficiently Using Prefix-Trees In Mining Frequent Itemsets", In Proceedings of the IEEE ICDM Workshop on Frequent Itemset Mining Implementations (FIMI'03), Melbourne, FL. B. Goethals and M. J. Zaki, Eds. Vol. 90. IEEE Press.

[34] Arun K Pujari, 2009 "Data Mining Techniques", Universities Press(India).

[35] Wojciechowski. M and Zakrzewicz.M, 2002 "Dataset Filtering Techniques in Constraint-Based Frequent Pattern Mining", Lecture Notes in Computer Science, Volume 2447.

[36] Tien Dung Do, Siu Cheung Hui and Alvis Fong, 2003 "Mining Frequent Itemsets with Category- Based Constraints", Lecture Notes in Computer Science, Volume 2843, pp. 76 – 86.

[37] Han. J. and Kamber. M, 2000 "Data Mining Concepts and Techniques", Morgan Kaufmann publishers.

[38] Han. J, 1995." Mining Knowledge At Multiple Concept Levels", In CIKM, 19-24.

[39] Han. J and Fu. Y, 1995 " Discovery Of Multiple-Level Association Rules From Large Databases", In Proc. of 1995 International Conference. on Very Large Data Bases (VLDB'95), ZAurich, Switzerland, September 1995, pp. 420-431.

[40] Psaila G and Lanzi P. L, 2000 " Hierarchy-based mining of association rules in data warehouses", In Proceedings of the 2000 ACM symposium on Applied computing 2000 , ACM Press, pp. 307-312.

[41] Yinbo WAN, Yong LIANG and Liya DING, 2009 "Mining Multi-level Association Rules From Primitive Frequent Item-sets", Journal of Macau University of Science and Tech-nology, June 30 , Vol 3 No 1.

[42] Brin S, Motwani R and Silverstein C, May 1997 "Beyond Market Baskets: Generalizing Association Rules to Correlations", In Proceedings of ACM SIGMOD Conference, pp. 265-276.

[43] Savasere A, Omiecinski E, and Navathe S, 1998 "Mining For Strong Negative Associations In A Large Database Of Customer Transactions " , In Proceedings of ICDE, pp. 494–502.

[44] Yuan X, Buckles B, Yuan Z and Zhang J, 2002 "Mining Negative Association Rules.", In Proceedings, Of ISCC , pp. 623–629.

[45] Wu X, Zhang C and Zhang S, 2002 " Mining Both Positive And Negative Association Rules", In Proceedings of ICML" pp. 658– 665.

[46] Wu X, Zhang C and Zhang S, 2004 " Efficient Mining of Both Positive and Negative Association Rules", ACM Transactions on Information Systems, Vol. 22, No. 3, pp. 381– 405.

[47] M.L. Antonie and O.R. Zaane, 2004 "Mining Positive and Negative Association Rules: an Approach for Confined Rules", Proceedings of International Conference on Principles and Practice of Knowledge Discovery in Databases, pp 27–38.

[48] Chris Cornelis, Peng Yan, Xing Zhang, and Guoqing Chen, 2006 " Mining Positive and Negative Association Rules from Large Databases", IEEE conference 2006.

[49] B.Ramasubbareddy, Dr.A.Govardhan, Dr.A.Ramamohanreddy, Nov 2010 " An Approach for Mining Positive and Negative Association Rules ", International Journal of Recent Trends in Engineering and Technology, Vol.4,No.1.

[50] Das. A, Ng, W.K and Woon Y.K, 2001. "Rapid Association Rule Mining", In Proceedings of the tenth international conference on Information and knowledge management, ACM Press, 474-481.

[51] M. Dunham, 2003 "D ata Mining Introductory and Advanced Topics", pp. 185-186, Section 6.7.2. Pearson Education.

[52] Bay Vo1 and Bac Le, Sep'2009 "Fast Algorithm for Mining Generalized Association Rules", International Journal of Database Theory and Application, Vol. 2, No. 3.

[53] JR.Srikant and R.Agrawal, 1995 "Mining Generalized Association rules", In Proceedings of the 21'st International Conference on Very Large Databases, pp.407-419, Zurich, Switzerland.

[54] Ashish Mangalampalli and Vikram Pudi, 2011 "Fuzzy Associative Rule-based Approach for Pattern Mining and Identification and Pattern-based classification", WWW 2011–Ph. D. Symposium .

[55] Ogunde A O, Folorunso O, Sodiya A S Oguntuase J A, and Ogunleye G O, 2011 " Improved Cost Models For Agent Based Association Rule Mining In Distributed Databases", Anale. Seria Informatica. Vol. IX.

[56] Christopher.T, 2010 " Character Based Weighted Support Threshold Algorithm Using Multi criteria Decision Making Technique ", International Journal On Computer science And Engineering Vol. 02, No. 04, 2010, pp. 965-971.

[57] A.Tiwari , R.K.Gupta and D.P.Agrawal , 2010 "A Survey On Frequent Pattern Mining:Current Status And Challenging Issues" Information Technology Journal 9(7) , pp. 1278-1293.

[58] B.Lent, A.Swami,and J.Wisdom, " Clustering association rules", In the proceeding of 13th International Conference on Data Engineering, pp.220.

[59] Rajendra K.Gupta and Dev Prakash Agarwal , "Improving the performance of Association Rule Mining Algorithms by Filtering Insignificant Transactions dynamically", Asian Journal of Information Management, pp.7-17. 2009 Academic Journals Inc.

[60] Pi Dechang and Qin Xiaolin," A New Fuzzy Clustering Algorithm on Association Rules for Knowledge

Management", Information Technology Journal. pp. 119-124, 2008. Asian Network for Scientific Information.

[61] Gyorodi C,Gyorodi R ,Coffey T,Holban S,2003 "Mining Association rules using Dynamic FP-trees, Proceedings of Irish signals and systems conference, Limerick, Ireland, 76-81.

[62] Claudia Marinica and Fabrice Guillet, June 2010, " Knowledge–Based Interactive Postmining of Association Rules Using Ontologies" IEEE Transactions On Knowledge And Data Engineering Vol.22 NO.6