International Journal of Computer Applications (0975 – 8887)
Volume 46– No.23, May 2012

A Survey on Frequent Itemset Mining with Association Rules

Endu Duneja
R.I.T.S, Bhopal
M.P., India

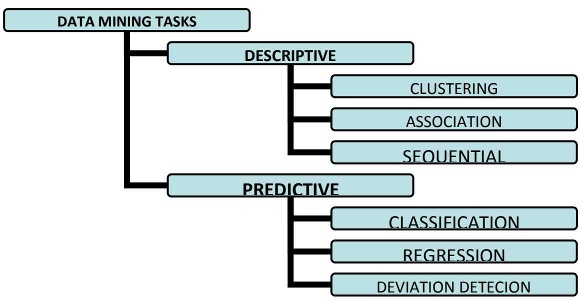
A.K. Sachan
R.I.T.S., Bhopal
M.P., India

ABSTRACT
Data mining techniques comprises of Clustering, Association, Sequential mining, Classification, Regression and Deviation detection Association Rule mining is one of the utmost ubiquitous data mining techniques which can be defined as extracting the interesting correlation and relation among huge amount of transactions. Many applications engender colossal amount of operational and behavioral data. Copious effective algorithms are proposed in the literature for mining frequent itemsets and association rules. Integrating efficacy considerations in data mining tasks is reaping popularity in recent years. Business value is enhanced by certain association rules and the data mining community has acknowledged the mining of these rules of interest since a long time. The discovery of frequent itemsets and association rules from transaction databases has aided many business applications. To discover the concealed knowledge from these data association rule mining can be applied in any application. A comprehensive analysis, survey and study of various approaches in existence for frequent itemset extraction, association rule mining with efficacy contemplations have been presented in this paper.

General Terms
Association Rule Mining, Frequent itemset mining.

Keywords
Frequent Itemset, Association Rule.

1. INTRODUCTION
Extracting Frequent Itemsets from transaction databases is a elementary task for several forms of knowledge discovery such as association rules, sequential patterns, and classification.



Association rule mining is an utmost ubiquitous data mining techniques, due to its substantial marketing custom and communities and many other diverse fields. Mining association rules significantly aid discovering relationships among items from large databases [2]. The source of motivation behind Association rule mining is the "market-basket analysis", a study on the habits of customers [3]. The mining of interesting associations, frequent patterns, correlations or casual structures among sets of items in the transaction databases or other data repositories is the main objective of Association rule mining [4].

The prompt enhancement of computer technology, specially augmented capacities and diminishing costs of storage media, has led businesses to store large amounts of outer and inner information in huge databases at minimal cost. Extracting useful information and beneficial knowledge from these huge databases has thus evolved into an important research area [5, 6, 7]. Among them association rule mining has been one of the utmost ubiquitous data-mining subjects, which can be defined as extracting interesting rules from large collections of data. Association rule mining has an extensive of applicability like Market basket analysis, Medical diagnosis/ research, Website navigation analysis, Homeland security and so on.

Association rules are used to find correlations among a set of items in database. These relationships are based on co-occurrence of the data items rather than inherent properties of the data themselves (as with functional dependencies). Association rules, first introduced in [8], the ensuing paper [9] is considered as one of the most significant contributions to the subject. Apriori, its main algorithm, not only affected the association rule mining community, but it influenced other data mining fields as well.

Association rule and frequent itemset mining became an extensive researched area, and hence quicker and speedier algorithms have been presented. Many of them are Apriori based algorithms or Apriori variations. Those who customized Apriori as a basic search strategy, be likely to adapt the complete set of procedures and data structures as well [8, 10, 11, 12]. Successively the scheme of this important algorithm was not only used in basic association rules mining, but also in association rules maintenance [15,16,13,17] , sequential pattern mining [18], episode mining, functional dependency discovery & other data mining fields (hierarchical association rules [12,13,14]. Frequent pattern mining methods can also be prolonged for the solution of many other problems, such as ice berg cube computation and classification. Thus the effectual and proficient frequent pattern mining is a significant and fascinating research problem.

2. ASSOCIAITON RULE MINING
In this section we will introduce association rule mining problem in detail. Different concerns in Association Rule Mining (ARM) will be expounded together.

Association Rule problem was first of all stated in [3] by Agrawal that the conventional statement of association rule mining problem was discovering the interesting association or correlation relationships among a large set of data items.

Let $\mathbf{I} = \boldsymbol{i_1, i_2, \ldots i_m}$ be a set of items. Let D, the task relevant data, be a set of database transactions where each transaction

T is a set of items such that $\mathbf{T} \subseteq \mathbf{I}$. Each transaction is associated with an identifier, called TID. Let A be a set of items. A transaction T is said to contain A if and only if $\mathbf{A} \subseteq \mathbf{T}$. An association rule is an implication of the form $\mathbf{A} \Longrightarrow \mathbf{B}$, where $\mathbf{A} \subset \mathbf{T}$, $\mathbf{B} \subseteq \mathbf{T}$ and $\mathbf{A} \cap \mathbf{B} = \emptyset$. The rule $\mathbf{A} \Longrightarrow \mathbf{B}$ holds in the transaction set D with support **s**, where **s** is the percentage of transactions in D that contain $\mathbf{A} \cup \mathbf{B}$. The rule $\mathbf{A} \Longrightarrow \mathbf{B}$ has confidence **c** in the transaction set D if c is the percentage of transactions in **D** containing **A** which also contain **B**,

$$\mathbf{Support(A \Rightarrow B) = prob\{A \cup B\}}$$
$$\mathbf{Confidence\ (A \Rightarrow B) = prob\ \{B/A\}}$$

The support s, of an association rule is the ratio (in percent) of the transactions containing X∩Y to the total number of transactions analyzed, |R(t)|. If the support of an association rule is 15% then it means that 15% of the analyzed transactions contain X∩Y. Support is the statistical significance of an association rule. The association rules have the supports less than 5% would be considered not very important to profile a user's behaviour. While a high support is often desirable for association rules.

For a given number of transactions, confidence c, is the ratio (in percent) of the number of transactions that contain X∪Y to the number of transactions that contain X. Thus if we say an association rule has a confidence of 77%, it means that 77% of the transactions containing X also contain Y. The confidence of a rule indicates the degree of correlation in the dataset between X and Y. It is used as a measure of a rule's strength. Often a large confidence is required for association rules.

Association rule mining is a two-step process:

Step 1: Discover all frequent itemsets.

Step 2: Create strong association rules from the frequent itemsets.

Typically, association rules are considered interesting if they satisfy both a minimum support threshold and a minimum confidence threshold. Such thresholds can be set by users or domain experts.

Rules that satisfy both a minimum support threshold (min sup) and a minimum confidence threshold (min conf) are called strong association rules.

k-itemset, an itemset that contains *k* items. The set {computer, antivirus} is a 2-itemset. The occurrence frequency of an itemset is the number of transactions that contain the itemset. This is also known, simply, as the frequency or support count of the itemset. An itemset satisfies minimum support if the occurrence frequency of the itemset is greater than or equal to the product of min sup and the total number of transactions in **D**. If an itemset satisfies minimum support, then it is a frequent itemset. The set of frequent k-itemsets is commonly denoted by $\mathbf{L_K}$ [1].

3. FREQUENT ITEMSET MINING

3.1 Introduction

The task of frequent itemset mining was introduced by Agrawal et al. [3] in 1993. A set of items that appears at least in a pre-specified number of transactions is called frequent itemset. Frequent itemsets are typically utilized to engender association rules.

The task of frequent itemset mining is defined as follows: Let I be a set of items. A set $\mathbf{X = \{\mathit{i_1, i_2, \ldots i_k}\}} \subset$ I is called an itemset, or k- a k-itemset, if it contains k items. A transaction over I is a couple T = (tid, I) where tid is the transaction identifier and I is an itemset. A transaction T = (tid, I) is said to support an itemset X $\subset$ I, if X $\subseteq$ I .A transaction database D over I is a set of transactions over I. The support of an itemset X in D is the number of transactions in D that supports X: Support (X, D) = | {tid | (tid, I) $\in$D, X $\subseteq$ I }|, frequency of an item set X in D is the probability of X occurring in a transaction T $\in$ D

$$Frequency\ (X, D) = P(X) = \frac{Support\ (X,D)}{|D|}$$
:

Note that |D| = support ({ }, D). An item set is called frequent if its support is no less than a given absolute minimal support threshold σ_{abs}, with $0 \leq \sigma_{abs} \leq |D|$. The frequent item sets discovered does not reflect the impact of any other factor except frequency of the presence or absence of an item.

3.2 Frequent Itemset Mining Destitution

Due to broad applications in mining association rules, correlations, and graph pattern constraint based on frequent patterns, sequential patterns, and many other data mining tasks studies of Frequent Itemset (or pattern) Mining is conceded in the data mining field because of its. Proficient algorithms for mining frequent itemsets are pivotal for mining association rules and also for many other data mining tasks. The paramount challenge observed in frequent pattern mining is enormous number of result patterns. An exponentially large number of itemsets are generated, as the least threshold sink. Consequently, pruning insignificant patterns can be done efficiently in mining process and that becomes one of the focal matters in frequent pattern mining.

Subsequently, core goal is to improve the process of finding patterns which should be effectual, accessible and can reveal the significant patterns which can be used in various ways. Frequent pattern: a pattern (a set of items, subsequences, substructures, etc.) that ensues frequently in a data set. Itemsets that gratifies the minimum support are frequent itemsets. Those itemsets that are anticipated or expected to be large or frequent are candidate itemsets [1].

4. LITERATURE SURVEY

In 1993 Agrawal, Imielinski, Swami [19] put forward one step for man, which leads a giant leap for computer science applications offered an algorithm AIS forefather of the algorithms to generate the frequent itemsets & confident association rule. It contains two stages. The first phase constitutes the generation of the frequent itemsets are generated in the first stage and in the next stage confident and frequent association rules are generated.

In 1995 SETM (SET-oriented Mining of association rules) [20] was motivated by the desire to use SQL to compute large itemsets. It utilized only simple database primitives, viz. sorting and merge-scan join. It was easy, rapid and durable over the variety of parameter values. It proved that some aspects of data mining can be carried out by using general query languages such as SQL, instead of developing specialized black-box algorithms. The set-oriented feature of SETM eased the development of extensions Apriori.

In 1994-95 the above algorithms were enhanced by Agrawal et al [21, 22] by using the monotonicity property of the support of itemsets and the confidence of association rules.

In 1995 Park et al. [23] planned an optimization, called Direct Hashing and Pruning (DHP) aimed towards curbing the number of candidate itemsets. They gave DHP algorithm for proficient large itemset generation. The recommended algorithm has two main traits: one is proficient generation for large itemsets and other is operative diminution on transaction database size. DHP is awfully proficient for the generation of candidate set for large 2-itemsets, in orders of magnitude, lesser than that by prior methods; it is so by using the hash techniques thus resolving the operational bottleneck.

In 1996 Agrawal et al [21, 24] proposed that the finest features of the Apriori & AprioriTid algorithms can be combined into a hybrid algorithm, called AprioriHybrid. Scale up tests showed that AprioriHybrid scrabbles linearly with the number of transactions. In adjunct, the execution time fall a little as the number of items in the database upsurge. As the average transaction size upsurge (however sustaining the database size constant), the execution time upsurges only slowly.

In 1997 Brin et al [25] proposed the DIC algorithm that partitions the database into intervals of a fixed size so as to lessen the number of traversals through the database. They put forth an algorithm for finding large itemsets which practices scarcer passes over the data than conventional algorithms, and yet uses scarcer candidate itemsets than approaches that rely on sampling. In addendum they have put forth a new way of spawning "implication rules", which are standardized based on both the predecessor and the successor. They bring into being more instinctive outcomes as compared to the antecedents.

In 1999 C. Hidber [26] put forth Continuous Association Rule Mining Algorithm (CARMA) used to figure large itemsets online. It incessantly generated large itemsets along with a dwindling support interval for each itemset. It exercise an identical technique in order to restrict the interval size to 1. C. He proved that CARMA's itemset lattice swiftly reckons a superset of all large itemsets while the sizes of the relating support intervals dwindle swiftly.

In 1999 Mohammed J. Zaki et al. [27] proposed Closed Association Rule Mining; (CHARM, 'H' is complimentary), an efficacious algorithm for mining all frequent closed itemsets. By using a dual itemset-Tidset search tree it reckoned closed sets, and also skive off many search levels by a proficient hybrid. Moreover using renowned diffsets technique it reduced the memory footmark of transitional computations. CHARM drastically outpaces erstwhile methods as proved by experimental assessment on a numerous real and imitate databases.

In 2000 M. J. Zaki [28] put forth new algorithms for detecting the set of frequent itemsets. Moreover he bestowed a lattice-theoretic methodology to partition the frequent itemset search space into trifling, autonomous sub-spaces by either maximal-clique-based or prefix-based approach. The valued results disclosed that the maximal-clique based decomposition is more scrupulous and directs to trifle classes.

In 2000 novel class of stimulating problem termed as WAR (weighted association rule) problem was ascertained by Wei Wang, et al. [29]. They put forth a line of attack for WARs by first snubbing the weight and discovering the frequent itemsets pursued by instituting the weight in the course of rule generation. The domino effect of the methodology is squatter average execution times, high quality domino effect fabrication too in comparison of generalization of prior methods on quantitative association rules.

Rivaling the Apriori [4] and its variants it is found that it need several database scans. Thus, in 2004 Jiawei Han et al. [32] proposed a new data structure, frequent pattern tree (FP-tree), for storing compressed, vital information about frequent patterns, and ripened a pattern growth method, this method needs only two database scans when mining all frequent itemsets for effectual mining. As the itemset in any transaction is always encoded in the corresponding path of the FP-trees consequently this method assured that it under no circumstances generates any combinations of new candidate sets which are absent in the database.

Since the FP-growth method [32] is brisker than the Apriori , it is found that few lately frequent pattern mining methods being effectual and scalable for mining long and short frequent patterns.

A conditional FP-tree is in orders of magnitude smaller rivaled to the global FP-tree. Consequently the size of the FP-trees to be handled would be considerably dwindled when a conditional FP-tree is created out of each projected database. This has been proved to be quicker than the Tree-Projection algorithm [16] where in the database is projected recursively into a tree of databases.

In addition to a randomized sampling-based algorithm and techniques for extending from frequent items to frequent itemsets in 2002 Manku and Motwani proposed Lossy Counting algorithm [33]. Their algorithm stores tuples which consists of an item, a lower bound on its count, and a 'delta' (Δ) value that records the difference between the upper bound and the lower bound.

In 2003 Mohammed J. Zaki et al. [34] have proposed a new vertical data depiction called Diffset which sustain trail of differences in the tids of a candidate pattern from its generating frequent patterns. They have proved diffsets drastically expurgated (by orders of magnitude) the extent of memory needed for keeping intermediate results. The execution time of vertical algorithms like Eclat [28] and CHARM [27] were convalesced by several orders of magnitude with abet of Diffsets.

In 2003 Ferenc Bodon [35] has explored theoretically and experimentally Apriori [21], is the most conventional algorithm for frequent itemset mining. The enactments of the Apriori algorithm have demonstrated considerable differences in execution time and memory requirement. He has amended Apriori and termed it as Apriori_Brave that seems to be swifter than the exemplar algorithm.

In 2003 an augmentation with a memory proficient data structure of a reckonable method to excavate association rules from data was recommended by Liang Dong et al. [36]. The best features of the Quantitative Approach, DHP, and Apriori were combined to constitute the recommended approach. The achieved outcomes precisely revealed the knowledge concealed in the datasets under investigation.

In 2004 a novel single-pass algorithm, termed as DSM-FI (Data Stream Mining for Frequent Itemsets) was implemented by Hua-Fu Li, et al. [37], which excavate all frequent itemsets over the history of data streams. DSM-FI outpace the Lossy Counting [33] in terms of running time and memory utilization between the large datasets.

In 2004 PRICES a proficient algorithm for excavating association rules was applied by Chuan Wang [38], which first recognizes all large itemsets and then forms association rules. His methodology decreased large itemset generation time, distinguished to be the most arduous step, by skimming the database just once and by logical operations in the process. It is capable and efficient and is ten times as quick as Apriori in some cases.

In 2005 M.H. Marghany & A.A. Mitwaly [39] proposed a fast algorithm for mining association rules using data structure in Java language known as Tree Map which store candidate as well as frequent itemsets & also used 'Array list' technique to reduce the need to traverse the database. The frequent itemsets are generated in alphabetical order which makes it simple for the user to discover the rules on as precise product. Their experimental outcomes outshine all prevailing algorithms in all common data mining problems.

In 2006 Mingjun Song et al. [40] have proposed a new operational algorithm for excavating complete frequent itemsets. They transmuted transaction ids of each itemset and compressed to continuous transaction gap lists in a separate space by the transaction tree and frequent itemsets were located by transaction gap connection along a lexicographic tree [30] in depth first order. The intersection time is significantly saved by compression. Their concept has outdone FPgrowth [31] and Eclat [34] on the basis of execution time and storage cost.

In 2006 Yanbin Ye et al. [41] have instigated a parallel Apriori algorithm based on Bodon's work [35] and scrutinized its enactment on a parallel computer. They followed a partition based Apriori algorithm to partition a transaction data set. They revealed that by fitting every partition into inadequate main memory for fast access and permitting incremental generation of frequent itemsets enhanced the performance of frequent itemsets mining.

In 2006 an operational and proficient Fuzzy Healthy Association Rule Mining Algorithm (FHARM) [42] has been applied by M. Sulaiman Khan, et al. In their methodology, comestible attributes were sifted from transactional input data by prognoses and then transfigured to RDA (Required Daily Allowance) numeric values. The averaged RDA database was then to transfigure a fuzzy database that contains normalized fuzzy attributes containing different fuzzy sets. By introducing new quality measures algorithm [47] fabricated more appealing and eminence rules.

In 2007 M. Hahsler, C. Buchta and K. Hornik gave a novel approach [44] based on the notion to firstly define a set of "interesting" itemsets and then, selectively generate rules for only these itemsets. The major benefit of this idea over swelling thresholds or filtering rules is that the number of rules found is considerably reduced whereas at the same time it is not obligatory to increase the support and confidence thresholds which may possibly lead to omitting significant information in the database.

In 2008 Kamrul et al [45] presented a novel algorithm Reverse Apriori Frequent pattern mining, which is a new methodology for frequent pattern production of association rule mining. This algorithm works proficiently, when the numerous items in the enormous frequent itemsets is near to the number of total attributes in the dataset, or if number of items in the hefty frequent itemsets is predetermined.

In 2008 E. Ansari et al [46] implemented DTFIM (Distributed Trie-based Frequent Itemset Mining). They used the Bodon's concept to distributed computing in a no shared memory multi computer environment to design their algorithm. They also added the concept from FDM for candidate generation step. The experimental results show that Trie data structure can be used for distributed association rule mining not just for sequential algorithms.

In 2010 S. Prakash & R.M.S. Parvathi [47] enhanced scaling Apriori for association rule mining efficacy by defining a new protocol suite that is the informative protocol suite that gives prediction sequences equal to those given by the association rule set by the confidence priority. The informative protocol suite is substantially smaller than the association rule set, specifically in case of smaller the minsup.

In 2012 Sanjeev Rao, Priyanka Gupta [48] proposed a novel scheme for mining association rules pondering the number of database scans, memory consumption, the time and the interestingness of the rules. They removed the disadvantages of APRIORI algorithm by determining a FIS data extracting association algorithm which is proficient in terms of number of database scan and time. They eradicate the expensive step candidate generation and also avoid skimming the database over and again. Thus they used Frequent Pattern (FP) Growth ARM algorithm that is more effectual structure to extract patterns when database intensifies.

5. PERFORMANCE ANALYSIS

In 1995 SETM (SET-oriented Mining of association rules) [20] outperformed AIS. Its set-oriented feature eased the development of extensions. The performance of AprioriTid degraded twice slow when applied to large problems however it accomplished equally well as Apriori for smaller problem sizes. Apriori and AprioriTid algorithms have always outdone AIS and SETM. There was substantial proliferation in the performance gap with the upsurge in problem size, ranging from a factor of three for small problems to more than an order of magnitude for large ones. Apriori outdone AIS on various problem size. It trounces by a factor of two for high minimum support and more than an order magnitude for low levels of support.

CARMA beats Apriori and DIC (Dynamic Itemset Counting) [25] on low support thresholds, also sixty times more memory efficient. The memory efficacy of CARMA was an order of magnitude greater than Apriori. CARMA is more effective on low support thresholds than Apriori and DIC [25]. Besides, the CARMA has been found to be sixty times more memory dexterous.

Symptomatically Pascal is twice as quick as A-Close [50], and ten times quicker than Apriori. A massive experimental estimation on a number of real and synthetic databases shows that CHARM appreciably outdone previous methods. CHARM can execute on very low support values [49] better than Pascal [49] besides being several orders of magnitude. CHARM outperforms Closet [51] by an order of magnitude or more, particularly in case of lowered support. The execution time of vertical algorithms like Eclat [28] and CHARM [27] were enriched by several orders of magnitude with the support of Diffsets. Tidset based methods are outperformed in several orders of magnitude by diffset algorithms. The average diffset size corresponding to long patterns is several orders of magnitude smaller than the analogous average Tidset size i.e. on dense sets, it is four to five orders of magnitude smaller whereas by only two to three orders for sparse sets.

Rivaled with Apriori [4] and its variants which need several database scans, the FP-growth method [30] require only two database skims when mining all frequent itemsets. It ascertained faster than the Tree-Projection algorithm [16] where in the database is projected recursively into a tree of databases.

The original algorithm [4] is outdone by Apriori_Brave [35]. The obtained results [24] precisely revealed the knowledge conceal in the datasets under scanning.

DSM-FI outpaces the Lossy Counting [33] in stipulations of running time and memory usage between the large datasets.

PRICES [38] is adept and dexterous and can sometimes be ten times as fast as Apriori.

The algorithm [38] has outdone FPgrowth [30] and Eclat [28] in terms of execution time and storage expense.

The classical and fuzzy ARM was used to resolve the problem of abrogation of DCP in weighted ARM. The issue of invalidation of downward closure property was resolved by using enhanced model of weighted support and confidence framework for classical and fuzzy association rule mining.

Algorithm	Apriori Hybrid [21]	DSM FI [37]	Parallel Apriori [41]	WAR [29]	Apriori Brave [35]	PRICE [38]	Combin ed [36]
Data Size (K)	100	2000	100	1000	100	100	100
Transaction Size	10	10	10	10	40	10	10
Itemset Size	4	5	4	4	10	4	4
Threshold	0.75	0.01	0.005	0.1	0.05	5	4
Time	7.5	400	998	1000	8.3	150	1500

Fig 1: Features of Various Algorithms

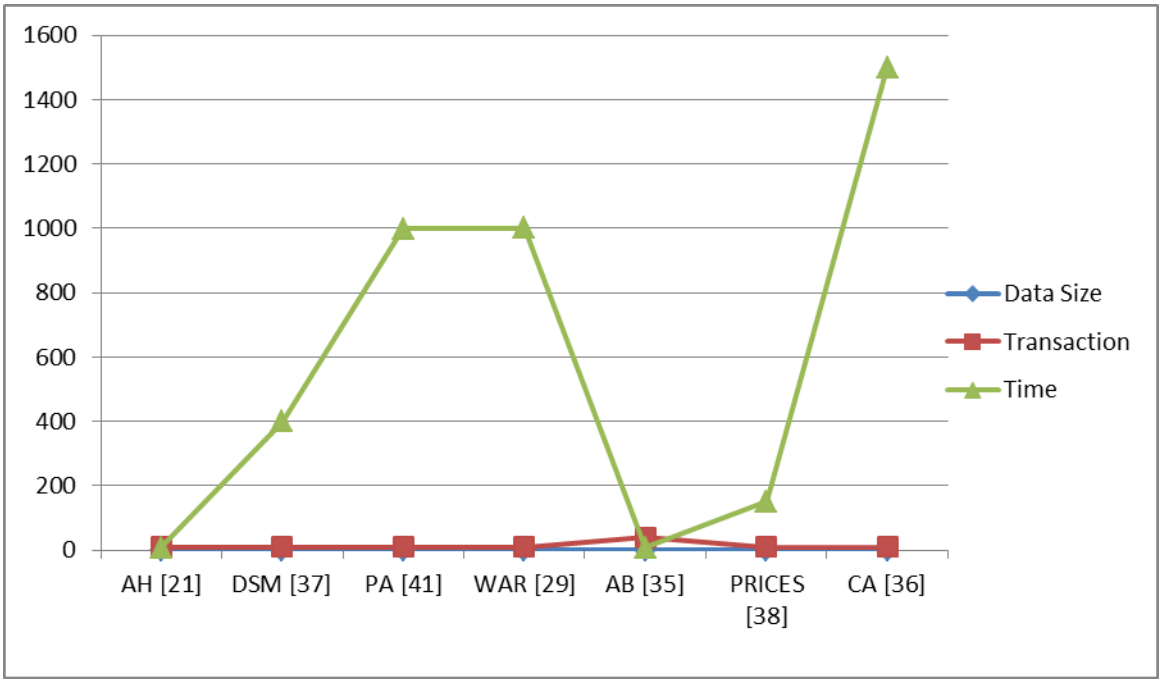


Fig 2: Performance Graph of Substantial Algorithms

The efficacy improvement outcomes from that the generation of the informative rule set [46] requires less candidates and database accesses than that of the association rule set rather than large memory usage like some other algorithms. The number of database accesses of the suggested algorithm is considerably lesser than other regular methods for generating association rules on all items.

6. CONCLUSION

Data mining imperative tasks are Association Rule mining & frequent itemset mining. Number of effective algorithms exists in the literature for extracting frequent itemsets & association rules. Integrating efficacy considerations in data mining tasks is achieving acceptance in past few years. The much admired fact of data mining community is discovering association rules used to erect the business of an enterprise. In this paper, we have accomplished a comprehensive survey of the algorithms and techniques in existence for the extracting frequent itemsets and association rules with efficacy considerations. An ephemeral analysis of numerous of algorithms is presented along with a relative study of a few substantial on the basis of their efficacy and memory expense.

7. ACKNOWLEDGMENTS

Authors are greatly thankful to Prof. Ashish Khare, Head of Department, at R.I.T.S, Bhopal, Dept. of Computer Science and Engg., for giving his valuable time and resources that helped me to do my research work. Also, we would like to thank our faculty members and Radharaman Group of Institutions, Bhopal, India for providing valuable suggestions in this research work.

8. REFERENCES

[1] Jiawei Han & Michelien Kamber, Edition 2003. "Data Mining: Concepts & Techniques", published by Elsevier, ISBN: 81-8147-049-4, pp 226-230.

[2] Yu-Chiang Li, Jieh-Shan Yeh, Chin-Chen Chang, 2005. "Efficient Algorithms for Mining Share-Frequent Itemsets", In Proceedings of the 11th World Congress of Intl. Fuzzy Systems Association.

[3] R.Agrawal, T.Imielinski, and A.Swami, 1993. "Mining association rules between sets of items in large databases", in proceedings of the ACM SIGMOD Int'l Conf. on Management of data, pp. 207-216.

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

[5] R.Agrawal, T. Imielinksi and A. Swami, Database Mining: a performance perspective, IEE Transactions on knowledge and Data Engineering, 1993.

[6] M.S. Chen, J.Han and P.S. Yu., 1996. "Data Mining : An overview from a database perspective", IEE Transactions on Knowledge and Data Engineering.

[7] C.Y. Wang, T.P. Hong and S.S. Tseng, 2002. "Maintenance of discovered sequential patterns for record deletion" in Intell. Data Anal., pp. 399-410.

[8] R.Agrawal, T. Imielinski, and A.Sawmi, 1993. "Mining association rules between sets of items in large databases", In proc. of the ACM SIGMOD Conference on Management of Data, pp. 207-216.

[9] R. Agrawal and R.Srikant.1994. "Fast algorithms for mining association rules", In Proc. of Intl. Conf. On Very Large Databases (VLDB), pp.487-499.

[10] S. Brin, R. Motwani, J.D.Vllman, and S.Tsur, 1997. "Dynamic itemset counting and implication rules for market basket data", SIGMOD Record (ACM Special Interest Group on Management of Data), vol. 26(2), pp.255.

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

[12] A.Sarasere, E.Omiecinsky,and S.Navathe, 1995. "An efficient algorithm for mining association rules in large databases" In Proc. 21St International Conference on Very Large Databases (VLDB) , Zurich, Switzerland, Technical Report No. GIT-CC-95-04.

[13] Y.F.Jiawei Han., 1995. "Discovery of multiple-level association rules from large databases" In Proc. of the 21St International Conference on Very Large Databases (VLDB), Zurich, Switzerland.

[14] Y.Fu., 1996. "Discovery of multiple-level rules from large databases".

[15] D.W.L. Cheung, J. Han and C.Y. Wong., 1996. "Maintenance of discovered association rules in large databases: An incremental updating technique" In ICDE, pp. 106-114.

[16] D.W.L.Cheung, S.D.Lee, and B.Kao., 1997. "A general incremental technique for maintaining discovered association rules" In Database Systems for advanced Applications, pp. 185-194.

[17] S.Thomas, S.Bodadola, K.Alsabti, and S.Ranka, 1997. "An efficient algorithm for incremental updation of association rules in large databases" In Proc. KDD'97, pp. 263-266.

[18] R. Agrawal and R. Srikant, 1995. "Mining sequential patterns" In P.S.Yu and A.L.P. Chen, editors, Proc.11the Int. Conf. Data engineering. ICDE, pp. 3-14. IEEE pp, 6-10.

[19] R.Agrawal, T.Imielinski, and A.Swami, 1993. "Mining association rules between sets of items in large databases", in proceedings of the ACM SIGMOD Int'l Conf. on Management of data, pp. 207-216.

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

[21] Rakesh Agrawal, and Ramakrishnan Srikant, 1994. "Fast Algorithms for Mining Association Rules", In Proceedings of the 20th Int. Conf. Very Large Data Bases, pp. 487-499.

[22] Srikant R, Agrawal R., 1995. "Mining generalized association rules". In: Dayal U, Gray P M D, Nishio Seds. Proceedings of the International Conference on Very Large Databases. San Francisco, CA: Morgan Kanfman Press, pp. 406-419

[23] J.S. Park, M.-S. Chen, and P.S. Yu, 1995. "An effective hash based algorithm for mining association rules", In Proceedings of the 1995 ACM SIGMOD International Conference on Management of Data, vol. 24(2) of SIGMOD Record, pp.175–186.

[24] R. Agrawal, H. Mannila, R. Srikant, H. Toivonen, and A.I. Verkamo, 1996. "Fast discovery of association rules" In Advances in Knowledge Discovery and Data Mining, MIT Press, pp. 307–328.

[25] 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, vol. 26(2), pp. 255–264.

[26] C. Hidber, 1999. "Online association rule mining". In Proc. of the 1999 ACM SIGMOD International Conference on Management of Data, vol. 28(2), pp. 145–156.

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

[28] M.J. Zaki, May/June 2000. "Scalable algorithms for association mining". IEEE Transactions on Knowledge and Data Engineering, vol.12(3), pp.372–390.

[29] Wei Wang, jiong yang and philip S. Yu, 2000. "Efficient Mining of Weighted Association Rules (WAR)", Proceedings of the sixth ACM SIGKDD international conference on Knowledge discovery and data mining, pp. 270 - 274.

[30] R. Agrawal, C. Aggarwal, and V. Prasad, 2000. "A Tree Projection Algorithm for Generation of Frequent Item Sets", Parallel and Distributed Computing, pp. 350-371.

[31] G. Grahne and J. Zhu, 2003. "Efficiently Using Prefix-Trees in Mining Frequent Itemsets", In proceedings of the IEEE ICDM Workshop on Frequent Itemset Mining Implementations (FIMI).

[32] Jiawei Han, Jian Pei, Yiwen Yin, Runying Mao, 2004. "Mining Frequent Patterns without Candidate Generation: A Frequent-Pattern Tree Approach", Data Mining and Knowledge Discovery, vol. 8, issue 1, pp. 53 – 87.

[33] G. S. Manku and R. Motwani, 2002. "Approximate Frequency Counts Over Data Streams". In Proc. of the 28th VLDB(Very Large Data Bases) conference, pp. 346–357.

[34] M.J. Zaki and K. Gouda, 2003. "Fast Vertical Mining Using Diffsets", Proc. Ninth ACM SIGKDD Int'l Conf. Knowledge Discovery and Data Mining, pp. 326-335.

[35] F. Bodon, 2003. "A Fast Apriori Implementation", In Proc. of the IEEE ICDM Workshop on Frequent Itemset Mining Implementations, vol. 90.

[36] Liang Dong and Christos Tjortjis, 2003. "Experiences of Using a Quantitative Approach for Mining Association Rules", in Lecture Notes Computer Science, vol. 2690, pp. 693-700.

[37] Hua-Fu Li, Suh-Yin Lee and Man-Kwan Shan, 2004. "An Efficient Algorithm for Mining Frequent Itemsets over the Entire History of Data Streams", In Proc. of the 1st Int'l. Workshop on Knowledge Discovery in Data Streams, pp. 20- 24.

[38] Chuan Wang, Christos Tjortjis, 2004. "PRICES: An efficient algorithm for mining association rules", in Lecture Notes Computer Science vol. 3177, pp. 352-358, ISSN: 0302-9743.

[39] M.H. Margahny & A.A. Mitwaly, 2005. "Fast algorithm for mining association rules", AIML Conference, CICC, Cairo, Egypt.

[40] Mingjun Song and Sanguthevar Rajasekaran, 2006. "A Transaction Mapping Algorithm for Frequent Itemsets Mining", IEEE Transactions on Knowledge and Data Engineering, vol. 18(4), pp. 472-481.

[41] Yanbin Ye and Chia-Chu Chiang, 2006. "A Parallel Apriori Algorithm for Frequent Itemsets Mining", Fourth International Conference on Software Engineering Research, Management and Applications, pp. 87- 94.

[42] M. Sulaiman Khan, Maybin Muyeba, Christos Tjortjis, Frans Coenen, 2006. "An effective Fuzzy Healthy Association Rule Mining Algorithm (FHARM)", In Lecture Notes Computer Science, vol. 4224, pp.1014-1022, ISSN: 0302-9743.

[43] M. Sulaiman Khan, Maybin Muyeba, Frans Coenen, 2008. "Fuzzy Weighted Association Rule Mining with Weighted Support and Confidence Framework", in ALSIP (PAKDD), pp. 52-64.

[44] M. Hahsler, C. Buchta and K. Hornik, 2007. "Selective Association Rule Generation", in Proceedings on Computational Statics, Springer.

[45] Kamrul, Shah, Mohammad, Khandakar, Hasnain, Abu, 2008. "Reverse Apriori Algorithm for Frequent Pattern Mining", Asian Journal of Information Technology, pp.524-530, ISSN: 1682-3915.

[46] E. Ansari, G.H. Dastghaibfard, M. Keshtkaran, H. Kaabi 2008. "Distributed Frequent Itemset Mining using Trie Data Structure", IAENG, vol.35:3.

[47] S. Praksh, R.M.S. Parvathi 2010. "An enhanced Scalling Apriori for Association Rule Mining Efficiency", European Journal of Scientific Research, vol. 39, pp.257-264, ISSN: 1450-216X.

[48] S. Rao, P. Gupta 2012. "Implementing improved algorithm over Apriori data mining association rule algorithm", IJCST, vol. 3, pp.489-493, ISSN: 2229-4333.

[49] Y. Bastide, R. Taouil, N. Pasquier, G. Stumme, and L. Lakhal, December 2000. "Mining frequent patterns with counting inference", SIGKDD Explorations, vol.2(2).

[50] N. Pasquier, Y. Bastide, R. Taouil, and L. Lakhal, January 1999. "Discovering frequent closed itemsets for association rules", In 7th Intl. Conf. on Database Theory.

[51] J. Pei, J. Han, and R. Mao,May 2000. "Closet: An efficient algorithm for mining frequent closed itemsets", In SIGMOD Int'l Workshop on Data Mining and Knowledge Discovery.