ISSN NO. 2320-5407

Journal homepage:http://www.journalijar.com

INTERNATIONAL JOURNAL OF ADVANCED RESEARCH

RESEARCH ARTICLE

National Symposium On Emerging Trends In Computing & Informatics, NSETCI 2016, 12th July 2016,Rajagiri School of Engineering & Technology, Cochin, India.

A Gleam of Light on Association Rule Mining and Reduction Techniques.

***Mini T V[1], R Nedunchezhian[2] and V Vijayakumar[3].**

1. Department of Computer Science,Sacred Heart College, Chalakudy.
2. Department of CSE, KIT–Kalaignarkarunanidhi Institute ofTechnology,Coimbatore.
3. Department of Computer Science,S.N.R. Sons College (Autonomous),Coimbatore.

Manuscript Info

Key words:
Data Mining.
Association Rule Mining (ARM)
Association Rule Reduction

Abstract

Association Rule Mining (ARM) algorithms generate an exceptionally large number of association rules, often in thousands or even millions. Further, the association rules are sometimes very extensive. The nearly impossible task of the end - users to comprehend or validate such large number of complex association rules and limits the usefulness of outcome of mining on massive data sources. Various strategies have been proposed to reduce the number of association rules. They include generating only “non-redundant” rules, generating only “interesting” rules, or generating only those rules satisfying certain other criteria such as coverage, leverage, lift or strength or pruning out ‘irrelevant’ rules. This paper presents the various methods used to solve the issue of association rule reduction.

Copy Right, IJAR, 2013,. All rights reserved.

Introduction:-

Association Rule Mining:Advancements in information technology have accelerated the collection, storage and processing of various sources of data in recent decades (Gupta &Nayak 2014). In the current era of information explosion, most of the established companies have accumulated enormous amount of databases with image, text, web and transactional data from their customers.

According to Fayyad &Uthurusamy (2002), this collection of transaction of data is expected to grow in an alarming fashion and the evolution of data mining technology is directly related to Moore's Law and Storage Law. Moore's Law states that computer processing power doubles every eighteen months. According to Storage Law, the disk storage capacity doubles in every nine months. This is precisely what has happened since 1960s, and the prediction correctly illustrates the stand point of mining technology today.

The aggressive growth rate of disk storage, Moore's Law and Storage Law growth and the gap between the trends represents a very interesting pattern in the state of technology evolution. This trend shows the ability to capture data from the far outpaced space to process and acquire meaningful knowledge from it. In the current era, the quantity of data available exceeds the analyzing capacity of the human beings. If the amount of information in the world doubles in every 18 months, the techniques that perform data analysis and interpretation for useful knowledge discovery and extraction become mandatory.

Since the beginning of the 90s, frequent pattern mining has become one of the most actively researched topics in data mining and knowledge discovery in databases (Aggarwal et al 2009). These research works analyze techniques which can compactly represent transaction databases which can be used for frequent pattern mining. Market basket analysis, web link analysis, click stream analysis, drug design (molecular fragment mining) and genome analysis are the various applications in which it is used.

For this task, a large number of efficient algorithms were developed, which are based on sophisticated data structures and clever processing schemes. Association rule mining is the most common approach for performing the task. The frequent pattern mining problem was first introduced by Agrawal et al (1993a, 1993b, 1993c) to mine association rules between sets of items. Frequent itemset mining aims at finding regularities in the shopping behaviour of customers of supermarkets, mail-order companies, on-line shops and so on and so forth. More precisely, they find sets of products that are frequently bought together.

Let $A = \{a_1, \ldots, a_m\}$ be a set of items. This set is called the item base. Items may be products, special equipments, service options and so on. Any subset $I \subseteq A$ is called an item set and an item set may be any set of products that can be bought together.

Let $T = (t_1, \ldots, t_n)$ with $\forall_i; 1 \leq i \leq n : t_i \subseteq A$ be a vector of transactions over A. This vector is called the transaction database. A transaction database can list, for example, the sets of products bought by the customers of a supermarket in a given period of time. Some general properties of a transaction database are

- Every transaction is an item set, but some item sets may not appear in T.
- Transactions need not be pair-wise different: it may be $t_i = t_j$ for $i \neq j$.
- T may also be defined as a bag or multi-set of transactions.
- The set A may not be explicitly given, but only implicitly as $A = \bigcup_{i=1}^{n} t_i$.

The first studied approach in the field of frequent pattern mining is based on association rules (Park et al 1995, Agrawal et al 1996). Association rules mining was first proposed to find all the rules in the transaction data to analyze the relationship between items purchased by customers in a shop. Association rule mining can be stated as follows.

Let $I = \{i_1, i_2, \ldots, i_m\}$ be a set of items. Let 'D' be a set of transactions, where each transaction 'T' is a set of items such that $T \subseteq I$. An association rule is an implication of the form, $X \rightarrow Y$, where $X \subset I$, $Y \subset I$ and $X \cap Y = \emptyset$. The rule $X \rightarrow Y$ holds in the transaction set T with confidence 'c', if c% of transactions in T that support 'X' also support 'Y'. The rule has support 's'inT if s% of the transactions in T contains $X \cup Y$.

A set of transactions D (the database) is given and the problem of mining association rule discovers all association rules and its support and confidence should be greater than the user-specified minimum support (called min_sup) and minimum confidence (called min_conf). An association rule 'r' is a relation between item sets of the form

$$r : X \Rightarrow (Y - X) \tag{1}$$

where X and Y are frequent itemsets and $X \subset Y$ is the itemsets of X and (Y – X) are called antecedent and consequence of the rule 'r' respectively.

A rule consists of an antecedent (left-hand side proposition) and consequent (right-hand side proposition) and it states that when the antecedent is true, then the consequent will also be true. Association rules are most frequently used for capturing such correlations. Each association rule is combined with two constraints namely, support and confidence, which are normally used to select interesting rules from the set of all possible rules. Support is defined as the proportion of transaction in the dataset which contains the itemset and confidence is defined as an estimate of the probability of finding the right-hand side of the rule in transactions under the condition that these transactions also contain the left-hand side.

The valid association rules are those for which the measure of support and confidence is greater than or equal to the minimal thresholds of support and confidence, called min_sup and min_conf (Webb 2000). Support and confidence are calculated as in Equations (2) and (3).

$$Support(X) = \frac{|\{t \in D \mid X \subseteq t\}|}{|D|} \tag{2}$$

$$\text{Confidence (r)} = \frac{\text{Support(Y- X)}}{Support(X)} \tag{3}$$

A typical association rule is of the form$A \Rightarrow B, C$ [Support = 60%, Confidence = 80%]
For example, the followinginteresting association rule can be generated from the web log access file."60% of visitors who accessed URLs B and C also visited A"
The process of finding all the association rules with support and confidence above the respective thresholds is a two stage process. The first stage consists of finding all sets of items with support above the support threshold; these sets are called large item sets. The second step consists of computing for each large itemsets confidence for all its expressions with the form of a rule; the expressions whose confidence is above the confidence threshold.

The pattern discovering task works with patterns (local structure in the data) and frequent patterns (patterns that occur frequently in a dataset). Much of data mining literature is concerned with formulating useful pattern structures and developing efficient algorithms for discovering frequent patterns. The importance of finding frequent patterns is two-fold. The first is that they can be used to discover useful rules and the second is that these discovered rules can then be used to discover some interesting regularities in the data. The main aim here is to find all useful pattern structures and use efficient algorithms to identify frequently occurring patterns. From these frequent patterns, useful rules can be generated, which can be used to infer knowledge.

The nearly impossible task of the end **-** users to comprehend or validate such large number of complex association rules and limits the usefulness of outcome of mining on massive data sources. Various strategies have been proposed to reduce the number of association rules. They include generating only "non-redundant" rules, generating only "interesting" rules, or generating only those rules satisfying certain other criteria such as coverage, leverage, lift or strength or pruning out 'irrelevant' rules. This paper presents the various methods used to solve the issue of association rule reduction.Usage of reduction techniques to improve the process of associative rule mining algorithm can be grouped into two approaches.

(i) Replacing the whole database with only part of it based on the current frequent itemsets or reducing the number of passes over the whole database.

(ii) To make the number of candidate itemsets much smaller, explore different kinds of pruning techniques

Both these techniques are motivated by the serious issue, that is, if the support and confidence thresholds are small, the set of association rules can develop to be unwieldy as the number of transactions increases. The number of rules presented to the user typically increases proportionately, as the number of frequent itemsets increases. The main objective of these algorithms includes reducing the huge set of frequent patterns generated while maintaining the high quality of patterns. This section presents the various methods used to solve this issue of associative rule mining.

Redundant Association Rules:-

Four types of research on mining association rules have been performed to address the problem of rule redundancy. Based on user-defined templates or item constraints first, the rules have been extracted (Baralis&Psaila 1997). Secondly, researchers have developed interesting measures to select only interesting rules (Hilderman& Hamilton 2002). Thirdly, to prune redundant rules and thus present smaller and regularly more comprehensible sets of association rules to the user (Cristofor&Simovici 2002), researchers have proposed inference rules or inference systems.

Another solution to the problem using the Maximum Entropy approach was presented by Jaroszewicz&Simovici (2002). For the most common cases the problem of efficiency of Maximum Entropy computations is addressed by using closed form solutions. Analytical and experimental evaluation specifies that it efficiently produces small sets of interesting association rules. A consistent need for human intervention in mining interesting association rules is necessary. This becomes most effective if the human analyst has a vigorous visualization tool for mining and visualizing association rules.Ashrafi et al (2004) presented several methods to eliminate redundant rules and to produce small number of rules from any given frequent or frequent closed itemsets generated. A three-step visualization method for mining market basket association rules was presented by Techapichetvanich&Datta (2004). Discovering frequent itemsets, mining association rules and finally visualizing the mined association rules are included in this step.

Additional redundant rule elimination methods are presented by Ashrafi et al (2005) that first identify the rules that have similar meaning and then eliminate those rules. Furthermore, their methods also help in removing redundant rules in such a way that they never drop any higher confidence or interesting rules from the resultant rule set.Ishibuchi& Yamamoto (2005) depicted how the rule weight of each fuzzy rule can be defined in fuzzy rule-based classification systems. As a first step, two heuristic methods for rule weight specification were proposed to improve classification.

Wu et al (2009) investigated an Improved Apriori-based Algorithm (IAA) for association rule mining where a novel count-based technique was used to extract the redundant candidate itemsets. A generation record was then used to reduce the scan time of database. Reduced Apriori Algorithm with Tag (RAAT) to reduce the number of frequent patterns generated by using tags was proposed by Yu et al (2008). This algorithm had the advantage of time efficiency by reducing the number of scans performed and it was later improved by Vyas et al (2010) by further extending tag parameter with JEP (Jumping Emerging Pattern).

Zhou et al (2010) proposed a multiple-segment algorithm that improved Apriori Algorithm using reduction schemes. First segment was focused on reducing the number of judgments passed during the generation of frequent candidate itemsets. Second segment stressed on trimming the frequent itemsets and the last segment used optimization techniques. In the same year, Aouad et al (2010) proposed a new distributed technique that considered inherent characters of Apriori algorithm to reduce the number of database scans.

Pancho et al (2013) introduced a novel approach called Fingrams to rule representation and simplification. In terms of co-fired rules, Fingrams depicts the graphical interaction between rules at the inference level and used a fuzzy based algorithm to detect redundancies and pruning.

Singh (2014) proposed an improved Apriori algorithm that reduced the scanning time. It was achieved by cutting down unnecessary transaction records as well as by reducing the redundant generation of sub-items during pruning the candidate itemsets. It can directly form the set of frequent itemsets and eliminate candidate having a subset that is not frequent.

Interestingness Measures:-

Correlations were used to identify both the absence and presence of items as a basis for generating the rules by Brin et al (1997). Chi-squared test for correlation from classical statistics was used to measure of significance of associations. The same authors used this test for support as part of their measure of interest of an association. However they used a metric called conviction as a measure of implication during rule generation. An approach to the rare item problem was presented by Liu et al (1999). The dilemma that arises in the rare item problem is the search for rules that involve infrequent (i.e., rare) items requiring a minimum support. But using this minimum support will typically generate many rules that are of no interest. Using a high support will eliminate the rules with rare items also reducing the number of rules mined. The authors overcame this problem by allowing users to specify different minimum supports for the various items in their mining algorithm.

Omiecinski (2003) focus on finding associations with a different slant, where a different view of significance was scrutinized. This method considered other measures, called all-confidence and bond instead of support. All these measures represented the degree to which items in an association are related to each other. If all rules that could be produced from that association have a confidence greater than or equal to a minimum all-confidence value, an association was deemed interesting. Here, if associations have a minimum all-confidence or minimum bond, then those associations will have a given lower bound on their minimum support and the rules produced from those associations will have a given lower bound on their minimum confidence as well.

The above work has similarities to the work of Ramaswamy et al (2008) except, data subsets are defined based on the data satisfying certain time constraints. The idea revolves around finding all itemsets that are frequent in a set of user-defined time intervals. In this case, the characteristics of the data define the subsets and not the end-user.

Negative Association Rules:-

Only the items enumerated in transaction are considered in typical association rules. Such rules are called positive association rules. Negative association rules also consider the same items, in spite of considering negated items (i.e. absent from transactions). Negative association rules are valuable in market-basket analysis to identify products that either conflict with each other or complement each other. Mining negative association rules is a complicated task, because there are essential differences between positive and negative association rule mining. The researchers are faced with two key problems in negative association rule mining:

(i) How to effectively search for interesting itemsets
(ii) How to effectively identify negative association rules of interest.
(iii)

Savasere et al (1998), the authors present an idea to mine strong negative rules. They join 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 dependent on the domain and requires a predefined taxonomy. A similar approach is described by Yuan et al (2002) and Wu et al (2004) derived an algorithm for generating both positive and negative association rules. They add on top of the support-confidence framework another measure called mininterestfor a better pruning of the frequent itemsets generated. Brin et al (1997) introduced the notion of negative relationships. Their model is chi-square based. They use the statistical test to verify the independence between two variables. To determine the nature (positive or negative) of the relationship, a correlation metric was used.

Pruning:-

Pruning methods for improving classification performance is one of the most extensively researched areas. Safavian and Landgrebe (1991), Kalles (1995), Breslow and Murthy (1998) and all these discuss different pruning strategies. In addition, empirical comparisons of a range of different pruning methods have also been conducted. Quinlan (1987) was one of the first to perform a comparison of pruning methods. He offered experimental results for three methods, namely, cost-complexity pruning, reduced error pruning and pessimistic error pruning. Subsequently, Mingers (1989) performed an experimental comparison of critical value pruning and minimum-error pruning along with the three pruning methods of Quinlan. Mingers neither used a common amount of data nor was the data set different for each of his experiments. To avoid this problem, Esposito et al (1997) used the same amount of data while generating the pruned decision tree.

Although all the above stated techniques have proved to be effective and achieved high accuracy, they contain some limitations. They performed well on relatively small datasets, but their performance degraded while applied with temporal databases and associative classifiers. Thus, separate studies were necessitated to analyze pruning methods for associative classifiers. Another issue according to Mutter et al (2004), despite associative classifiers producing accurate results, it also produces huge number of rules and therefore is slow.

Many studies (Agrawal&Srikant 1994, Han et al 2000a and 2000b) have indicated the inherent nature of a combinatorial explosive number of frequent patterns. Hence association rules could be generated when the support threshold is small. To achieve high accuracy, a classifier may have to handle a large set of rules. These include storing those generated by association mining methods, retrieving the related rules, and pruning and sorting a large number of rules.

Data mining algorithms involve the production of huge sets of rules. These are unworkable for the analysis to be done without automation to develop methods for removing redundant rules from those sets. This problem can be resolved by using the Maximum Entropy (ME) approach (Jaroszewicz&Simovici 2002). Using closed form solutions for the most frequent cases the problem of efficiency of ME computations is dealt with. Analysis and experimental assessment of the proposed approach demonstrates the effective construction of small sets of interesting association rules.

Chawla et al (2004) offered an adaptive technique for local pruning approach for association rules. This approach utilizes the precise mapping between a certain class of association rules, which consist of those whose consequents are singletons and backward directed hypergraphs. The hypergraph representing the association rules is called an Association Rules Network (ARN). It provides a mechanism for fusing association rules in structured manner. There are two operations on this network for pruning rules to prove several properties of the ARN and use the results obtained with our technique to two popular data sets. The pruning process is adaptive based on the choice of the goal node made by the user. In the hypergraph pruning is decreased to cycle and reverse edge detection.

Top-k closed frequent patterns (TFP) of length no less than min_k can be discovered forming top-k most frequent closed patterns (Wang et al 2005). TFP gradually raises the support threshold during the mining. It prunes the FP-tree both during and after the tree construction phase. The top-k most frequent patterns usually do not represent the most representative k patterns due to the uneven frequency distribution among itemsets, Another branch of the work takes a “summarization” approach which is aimed at deriving k representatives which cover the whole set of (closed) frequent itemsets. The k representatives give compact compression over the collection of frequent patterns, making it easier to interpret and use. A profile-based approach to summarize a set of (closed) frequent itemsets into k representatives was proposed by Yan et al (2005). A "profile" over a set of similar itemsets can be defined as a union of these itemsets, as well as item probability distribution in the supporting transactions. Profile-based approach highlights its ability in restoration of individual itemsets and their supports with small error.

A general model forming approximate frequent itemsets (AFI) was developed by Liu et al (2006) to control errors of two directions in matrices formed by transactions and items. Siebes et al (2006) proposed a formulation with the MDL principle ie., the best set of frequent itemsets is the set, that compresses the database best. For finding the subset of frequent itemsets that compresses the database, heuristic algorithms are developed. Real data is typically subject to noise and measurement error, so instead it is demonstrated through theoretical results that, in the presence of even low levels of noise, large frequent itemsets are broken into fragments of logarithmic size; in effect the itemsets cannot be recovered by a routine application of frequent itemset mining.

Chen et al (2011) described the basic ideas and the disadvantages of Apriori algorithm. The low performance and efficiency of the algorithm caused generation of lots of candidate sets, and scanning the transaction database repeatedly. As a solution of this problem, it studied the pruning optimization and transaction reduction strategies. The enhanced Apriori algorithm based on pruning optimization and transaction reduction has been used to reduce the running time and to enhance the performance. Pruning achieves both complexity reduction of the final hypothesis for improved understandability, and providing an improvement in predictive accuracy. It is accomplished by minimizing the disturbances due to noisy data. A new hybrid pruning approach for rule induction, as well as an incremental post pruning method based on a misclassification tolerance was introduced by Shehzad (2013). Both these techniques were designed for RULe Extraction System (RULES-7), but are also applicable to any rule induction algorithm.

Classification Based Association (CBA) classifier has also been used interchangeably for the initial pruning of rules. According to Domínguez et al (2014) pruning is based on a computation that assesses how the presence of an item in the antecedent affects the confidence of the rule. Several experiments were carried out to check the effect of this method of pruning. Pruning method has been defined to eliminate rules using the Confident Interval (CI) measure. This again performs the evaluation of the change in the confidence of a rule due to the availability of an item in the antecedent.

The following table summarises the issues and solutions of association rule reduction.

Table 1. Issues and solutions of association rule reduction

S. No	ARM Methods and Issues	Authors &Year	Solutions
1	Redundant Association Rules - Number of redundant rules is larger than number of essential rules	Ashrafi et al (2004)	several methods to eliminate redundant rules and to produce small number of rules
2		Ishibuchi& Yamamoto (2005)	Specification of rule weight
3		Yu et al (2008)	Reduced Apriori Algorithm with Tag (RAAT) to reduce the number of frequent patterns
4		Wu et al (2009)	Improved Apriori-based Algorithm (IAA) for reducing redundant rules
5		Zhou et al (2010)	multiple-segment algorithm that improved Apriori Algorithm
6		Pancho et al (2013)	Fingrams to rule representation
7		Singh (2014)	improved Apriori algorithm
8	Interestingness Measures	Omiecinski (2003)	all-confidence and bond introduced instead of support
9		Ramaswamy et al (2008)	data subsets are defined based on the data satisfying certain time constraints
10	Negative Association Rules - Search for interesting itemsets and identify negative association rules of interest	Savasere et al (1998)	join positive frequent itemsets with domain knowledge in the form of taxonomy to mine negative associations
11		Yuan et al (2002) and Wu et al (2004)	added mininterest with support-confidence framework
12	Pruning	Domínguez et al (2014)	Eliiminated rules using the Confident Interval (CI) measure
13		Shehzad (2013)	hybrid pruning approach for rule induction
14		Chen et al (2011)	enhanced Apriori algorithm based on pruning optimization and transaction reduction
15		Siebes et al (2006	frequent itemsets formulation with the MDL principle
16		Yan et al (2005)	profile-based approach
17		Wang et al 2005	Top-k closed frequent patterns
18		Chawla et al (2004)	adaptive technique for local pruning approach
19		Jaroszewicz&Simovici 2002	Maximum Entropy (ME) approach

Conclusion

This paper reviewed the various research works conducted in the field of associative rule mining. Although several algorithms have been developed the search for methods that makes the mining algorithms perform well with high dimensional temporal dataset is still very active. Association rule mining is one of the widely applied data mining techniques that search for valuable relationships among different items in a dataset. Association rule mining is to find relationship between items in an item domain. The knowledge discovered using ARM can be used in various application areas like business, scientific and engineering applications. This paper has conducted a review on various ARM algorithms and Association Rule reduction methods and the accuracy of the existing methods still leaves room for improvement in future.

References:-

1. **Aggarwal, C.Li., Wang, Y.J.and Wang, J.(2009).**Frequent pattern mining with uncertain data.Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM.29-38.
2. **Agrawal, R.,and Shafer, J.C.(1996).** Parallel mining of association rules: design, implementation and experience.IEEE Transactions Knowledge Data Engineering. 8(6): 962–969.
3. **Agrawal, R.and Srikant, R.(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, Santiago, Chile. 487-499.
4. **Agrawal, R., Imielinski, T.and Swami, A.(1993a).** Database mining: A performance perspective.IEEE Transactions on Knowledge and Data Engineering. 5(6): 914-925.

5. **Agrawal, R., Imielinski, T.and Swami, A.(1993b).** A Mining association rules between sets of items in large databases.Proc. ACM SIGMOD Conference on Management of Data. 207–216
6. **Agrawal, R., Imielinski, T.and Swami, A.(1993c).** Mining association rules between sets of items in large databases.Buneman, P. and Jajodia, S., editors, Proceedings of the ACMSIGMOD International Conference on Management of Data, ACM Press, Washington. 207-216
7. **Aouad, L.M., Le-Khac, N.A.and Kechadi, T.M.(2010).**Performance study of distributed Apriori-like frequent itemsets mining. Knowledge and information systems.23(1): 55-72.
8. **Ashrafi, M., Taniar, D.and Smith, K.(2004).**A New Approach of Eliminating Redundant Association Rules. Lecture Notes in Computer Science. 3180: 465– 474.
9. **Ashrafi, M., Taniar, D.and Smith, K.(2005).** Redundant Association Rules Reduction Techniques. Lecture Notes in Computer Science. 3809: 254–263.
10. **Baralis, E.and Psaila, G.(1997).**Designing templates for mining association rules.Journal of Intelligent Information Systems. 9(1): 7-32.
11. **Brin, S., Motwani, R.and Silverstein, C.(1997).** Beyond market basket: Generalizing association rules to correlations', Proceedings of the ACM SIGMOD '97 International Conference on Management of Data. 265–276.
12. **Brin, S., Motwani, R., Ullman, J.and Tsur, S.(1997).**Dynamic itemset counting and implication rules for market basket data.Proceedings of the International ACM SIGMOD Conference, Tucson, Arizona, USA.255-264.
13. **Chawla, S., Davis, J.G.andPandey, G.(2004).** On Local Pruning of Association Rules Using Directed Hypergraphs.In ICDE. 4: 832-841.
14. **Chen, Z., Cai, S., Song, Q.and Zhu, C.(2011).** An improved Apriori algorithm based on pruning optimization and transaction reduction In Artificial Intelligence.Management Science and Electronic Commerce (AIMSEC).1908-1911.
15. **Cristofor, L.and Simovici, D.(2002).**Generating an informative cover for association rules.Proceedings of the IEEE International Conference on Data Mining.1-8.
16. **Domínguez, O., Juan L., Victoria, P.Á.and Jacinto, M.V.(2014).** Improving the Classifier CBA with an Additional Pruning Method.The Third International Conference on Informatics Engineering and Information Science (ICIEIS2014), Lodz University of Technology, Lodz, Poland.91-95.
17. **Esposito, F., Malerba, D.and Semeraro, G.(1997).**A Comparative analysis of methods for pruning decision trees.IEEE Transactions on Pattern Analysis and Machine Intelligence. 19(5): 476-491.
18. **Fayyad, U.and Uthurusamy, R.(2002).**Evolving data into mining solutions for insights.Communications of ACM. 45(8): 28-31.
19. **Gupta, S.K.and Nayak, R.P.(2014).**Advances in collating and using trial data.Chevret, S. and Resche-Rigon, M. (Eds.), Future Science Ltd., London.
20. **Han, J., Pei, J.and Yin, Y.(2000a).**Mining frequent patterns without candidate generation.Proceeding of the 2000 ACM-SIGMOD International Conference on Management of Data (SIGMOD00) Dallas, TX.1-12.
21. **Han, J., Pei, J., Mortazavi-Asl, B., Chen, Q., Dayal, U.and Hsu, M.(2000b).**FreeSpan: Frequent pattern-projected sequential pattern mining.ACM-SIGKDD.355-359.
22. **Hilderman, R.J.and Hamilton, H.J.(2002).** Knowledge Discovery and Interest Measures, Kluwer Academic, Boston.
23. **Ishibuchi, H.and Yamamoto, T.(2005).**Rule weight specification in fuzzy rule-based classification systems.IEEE Transactions on Fuzzy Systems.13(4): 428-435.
24. **Jaroszewicz, S.and Simovici, D.A.(2002).** Pruning redundant association rules using maximum entropy principle.Advances in Knowledge Discovery and Data Mining , Springer Berlin Heidelberg. 2336: 135-147.
25. **Jaroszewicz, S.and Simovici, D.A.(2002).** Pruning redundant association rules using maximum entropy principle.Advances in Knowledge Discovery and Data Mining , Springer Berlin Heidelberg. 2336: 135-147.
26. **Kalles, D.(1995).**Decision trees and domain knowledge in pattern recognition.Ph.D Thesis, Department of Computation, UMIst.
27. **Liu, B., Hsu, W.and Ma, Y.(1999).** Mining association rules with multiple minimum supports.Proceedings of fifth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Diego, California.337-341.
28. **Liu, H., Han, J., Xin, D.and Shao, Z.(2006).** Mining frequent patterns on very high dimensional data: a top-down row enumeration approach. Proceeding of the 2006 SIAM International Conference on Data Mining (SDM06) Bethesda, MD. 280–291.

29. **Mingers, J.(1989).**An empirical comparison of pruning methods for decision tree induction.Machine Learning. 4(2): 227-243.
30. **Mutter, S., Hill, M.and Frank, E.(2004).**Using classification to evaluate the output of confidence-based associated rule mining.The 17th Annual Joint Conference on Artificial Intelligence, Cairns, University of Freiburg, Germany.
31. **Omiecinski, E.(2003).**Alternative Interest Measures for Mining Associations in Databases.IEEE Transactions on Knowledge and Data Engineering. 15(1): 57-69.
32. **Pancho, D.P., Alonso, J.M., Cordón, O., Quirin, A.and Magdalena.(2013).** FINGRAMS: visual representations of fuzzy rule-based inference for expert analysis of comprehensibility.IEEE Transactions on Fuzzy Systems.18-24.
33. **Park, J.S., Chen, M.S.and Yu, P.S.(1995a).**An effective hash-based algorithm for mining association rules.Proceeding of the 1995 ACM-SIGMOD International Conference on Management of Data (SIGMOD95) San Jose, CA. 175-186.
34. **Park, J.S.,Chen,M.S. and Yu, P.S.(1995b).**An efficient parallel data mining for association rules. Proceeding of the fourth international conference on information and knowledge management.31-36.
35. **Quinlan, J.(1987).** Simplifying decision trees, International Journal of Man-Machine Studies. 27(3): 221-234.
36. **Ramaswamy, S., Mahajan, S.and Silbershatz, A.(2008).**On the Discovery of Interesting Patterns in Association Rules.Proceedings of Very Large Databases Conference.368-379.
37. **Safavian, S.R.and Landgrebe, D.(1991).**A survey of decision tree classifier methodology.IEEE Transactions on Systems, Man, and Cybernetics. 21(3): 660-674.
38. **Savasere, A., Omiecinski, E.and Navathe, S.(1995).** An efficient algorithm for mining association rules in large databases.Proceeding of the 1995 International Conference on Very Large Databases (VLDB95) Zurich, Switzerland.432–443.
39. **Shehzad, K.(2013).**Simple hybrid and incremental postpruning techniques for rule induction. IEEE Transactions on Knowledge and Data Engineering. 25(2): 476-480
40. **Siebes, A., Vreeken, J.and Leeuwen, M.(2006).** Item sets that compress. Proceeding of the 2006 SIAM International Conference on Data Mining (SDM06) Bethesda, MD. 393–404.
41. **Singh, J., Ram, H.and Sodhi, D.J.(2014).**Improving efficiency of apriori algorithm using transaction reduction.International Journal of Scientific and Research Publications.3(1): 1-4.
42. **Techapichetvanich, K.andDatta.(2004).** Visual Mining of Market Basket Association Rules. Lecture Notes in Computer Science. 3046: 479–488.
43. **Vyas, Z.V., Ganatra, A.P., Kosta, Y.P.and Bhesadadia, C.L.(2010).**Modified RAAT (Reduced Apriori Algorithm using Tag) for efficiency improvement with EP (Emerging Patterns) and JEP (Jumping EP).International Conferenec on Advances in Computer Engineering.238-240.
44. **Wang, J., Han, J., Lu, Y.and Tzvetkov, P.(2005).** TFP: An efficient algorithm for mining top-k frequent closed itemsets.IEEE Transactions on Knowledge Data Engineering. 17: 652–664.
45. **Webb, G.I.(2000**).Efficient Search for Association Rules. R. Ramakrishnan and S. Stolfo (Eds.) Proceedings of the Sixth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD-2000) Boston, MA. New York: The Association for Computing Machinery. 99-107.
46. **Wu, X., Zhang, C.and Zhang, S.(2004).** 'Efficient Mining of Both Positive and Negative Association Rules', ACM Transactions on Information Systems. 22(3): 381–405.
47. **Wu, Z., Li, C., Lv, S.and Zhou, B.(2009).**Pantothenate kinase-associated neurodegeneration: insights from a Drosophila model. Hum. Mol. Genet. 18(19): 3659-3672.
48. **Yan, X., Yu, P.S.and Han, J.(2005).**Substructure similarity search in graph databases.Proceeding of the 2005 ACM-SIGMOD International Conference on Management of Data (SIGMOD05) Baltimore, MD.766–777.
49. **Yu, W., Wang, X., Wang, F., Wang, E.and Chen, B.(2008).** Notice of Retraction The research of improved apriori algorithm for mining association rules.11th IEEE International Conference on Communication Technology (ICCT).513-516.
50. **Yuan, X., Buckles, B., Yuan, Z.and Zhang, J.(2002).**Mining negative association rules.Proceedings of ISCC.623–629.
51. **Zhou, Y., Wan, W., Liu, J.and Cai, L.(2010)**. Mining association rules based on an improved Apriori Algorithm.International Conference on Audio Language and Image Processing (ICALIP). 414-418.