Attribute Relational Analysis (ARA) For Coherent Association Rules: A Post Mining Process For Parallel Edge Projection And Pruning (PEPP) Based Sequence Graph Protrude Approach For Closed Itemset

Kalli Srinivasa Nageswara Prasad[1] and Prof. S. Ramakrishna[2]

[1] Sri Venkateswara University, Tirupati

Received: 1 November 2011 Accepted: 21 November 2011 Published: 3 December 2011

Abstract

Association rules present one of the most impressive techniques for the analysis of attribute associations in a given dataset related to applications related to retail, bioinformatics, and sociology. In the area of data mining, the importance of the rule management in associating rule mining is rapidly growing. Usually, If datasets are large, the induced rules are large in volume. The density of the rule volume leads to the obtained knowledge hard to be understood and analyze. One better way of minimizing the rule set size is eliminating redundant rules from rule base. Many efforts have been made and various competent and excellent algorithms have been proposed. But all of these models relying either on closed itemset mining or expert?s evaluation. None of these models are proven best in all data set contexts. Closed itemset model is missing adaptability and expert?s evaluation process is resulting different significance for same rule under different expert?s view. To overcome these limits here we proposed a post mining process called ARA as an extension to our earlier proposed closed itemset mining algorithm called PEPP.

Index terms— post mining, association rule mining, closed itemset, PEPP, Inference analysis, rule pruning.

1 INTRODUCTION

n general, association rules tend to deliver an efficient method of analysing binary or discretized data sets that are large in volume. One common practice is to determine relationships between binary variables in transaction databases, which is known as 'market basket analyses. In the case of non-binary data, initially data being coded as binary and then association rules will be used to analyse. Association rules having their impact on analysing large binary datasets and considered as versatile approach for modern applications such as detection of bio-terrorist attacks [1] and the analysis of gene expression data [2], to the analysis of Irish third level education applications [4].

The steps involved in a typical association rule analysis are "Coding of data as binary if data is not binary" -> "Rule generation" -> "Post-mining". This survey focused on post mining. It was a century after the introduction of association rules (associations initially discussed in 1902), it is still continuing that the absence of items from transactions is often ignored in analyses. a) Association rule mining Given a set I that is non-empty, a rule of association is a statement of the form X ? Y, where X, Y ? I such that X ? ?, Y ? ?, and X ? Y = ?. The dataset X is called the antecedent of the rule, the set Y is called the consequent of the rule, and we shall call I the master itemset. Association rules are generated over a large set of transactions, denoted by T. Xn association rule can be deemed interesting if the items involved occur together often and there are suggestions that one of the sets might in some cases lead to the presence of the other set. Xn association rules are characterised as interesting, or not, based on mathematical notions called 'support', 'confidence' and 'lift'. Although there are now a multitude of measures of interestingness available to the analyst, many of them are still based on these three functions.

In many applications, it is not only the presence of items in the antecedent and the consequent parts of an association rule that may be of interest. Consideration, in many cases, should be given to the relationship between the absence of items from the antecedent part and the presence or absence of items from the consequent part. Further, the presence of items in the antecedent part can be related to the absence of items from the consequent part; for example, a rule such as {margarine} ? {not butter}, which might be referred to as a 'replacement rule'. One way to incorporate the absence of items into the association rule mining paradigm is to consider rules of the form X?/ Y [20]. Another is to think in terms of negations. Suppose X ? I , then write ¬X to denote the absence or negation of the item, or items, in X from a transaction. Abstract -Association rules present one of the most impressive techniques for the analysis of attribute associations in a given dataset related to applications related to retail, bioinformatics, and sociology. In the area of data mining, the importance of the rule management in associating rule mining is rapidly growing. Usually, If datasets are large, the induced rules are large in volume. The density of the rule volume leads to the obtained knowledge hard to be understood and analyze. One better way of minimizing the rule set size is eliminating redundant rules from rule base. Many efforts have been made and various competent and excellent algorithms have been proposed. But all of these models relying either on closed itemset mining or expert's evaluation. None of these models are proven best in all data set contexts. Closed itemset model is missing adaptability and expert's evaluation process is resulting different significance for same rule under different expert's view. To overcome these limits here we proposed a post mining process called ARA as an extension to our earlier proposed closed itemset mining algorithm called PEPP.

2 Global

3 I)

2011 December (of the items in X is equivalent to X = 1, while the negation ¬X is equivalent to XX'?=X'?0. The concept of considering association rules involving negations, or "negative implications", is due to Silverstein et al [20].

4 b) Post Mining

Pruning rules and detection of rule interestingness are employed in the post-mining stage of the association rule mining paradigm. However, there are a host of other techniques used in the post-mining stage that do not naturally fall under either of these headings. Some such techniques are in the area of redundancy-removal. There are often a huge number of association rules to contend with in the post-mining stage and it can be very difficult for the user to identify which ones are interesting. Therefore, it is important to remove insignificant rules, prune redundancy and do further post-mining on the discovered rules **??**18,11]. Liu et al [11] proposed a technique for pruning and summarising discovered association rules by first removing those insignificant associations and then forming direction-setting rules, which provide a sort of summary of the discovered association rules. Lent et al. [19] proposed clustering the discovered association rules.

5 c) Influences of Input formats

The input formats that influence the post mining methodologies are binary data, text data and streaming data. In association rule mining, much of recent research work has focused on the difficult problem of mining data streams such as click stream analysis, intrusion detection, and web-purchase recommendation systems. In the case of streaming data, it is not possible to perform mining on cached and fixed data records. The attempt of caching data leads to memory usage issues, and the attempt of mining static data leads to worst time complexity since continuous dataset update leads to continuous passes through dataset. From the data mining point of view, texts are complex data giving raise to interesting challenges. First, texts may be considered as weakly structured, compared with databases that rely on a predefined schema. Moreover, texts are written in natural language, carrying out implicit knowledge, and ambiguities. Hence, the representation of the content of a text is often only partial and possibly noisy. One solution for handling a text or a collection of texts in a satisfying way is to take advantage of a knowledge model of the domain of the texts, for guiding the extraction of knowledge units from the texts. One of the obvious hot topics of data mining research in the last few years has been rule discovery from binary data. It concerns the discovery of set of attributes from large binary records such that these attributes are true within a same line often enough. It is then easy to derive rules that describe the data, the popular association rules though the interest of frequent sets goes further.

Based on the proposals [14,12,10,3] recently cited in literature and their motivations, it is observable that the process of rule pruning will opt to one of the two models.

1) Rule pruning under post mining process that demands domain experts observation 2) Rule pruning under post mining process that aims to avoid domain experts role in pruning process.

As in the first case rule pruning accuracy depends on domain expert's awareness on attribute relations. In this case it is always obvious to prune the rules under reliable domain expert's observation. In second case the models prune the rules based on dynamically determined attribute relations. This limits the solution to specific data models. Hence it is not adaptable for all contexts.

The Rest of the paper organized as; in section II we discussed the most frequently cited post mining models to improve rule accuracy. Section III briefs the post mining process [] that we opted. Section IV briefs the approach of closed itemset mining, and inference approach for itemset pruning. Section V explores the process of Attribute Relational weights analysis for rule pruning. Results discussion and comparative study will be in Section VI that fallowed by conclusion and references.

6 II.

7 RELATED WORK

Huawen Liu, et al [14] proposed post processing approach for rule reduction using closed set to filter superfluous rules from knowledge base in a post-processing manner that can be well discovered by a closed set mining technique. Most of these methods are based on the rule structure analysis where the relations between the rules have been analysed using corresponding problem. This procedure claims to eliminate noise and redundant rules in order to provide users with compact and precise knowledge derived from databases by data mining method. Further, a fast rule reduction algorithm using closed set is introduced. Other endeavours have been attempted to prune the rule bases directly. The typical cases have been elaborated and illustrated by eminent people from all over.

Modest number of proposals is addressed on pre-pruning and post-pruning. In a line, the pruning operation occurs at the phase of generation of significant rules. To add to this, the post pruning technique mainly concerns primarily emphasises that pruning operation occurs after the rule generation, among which the rule cover is a representative case. To extract interesting rules, approri knowledge has also been taken into account in literatures and a template An association rule is an implication expression illustrates a kind of association relation. A rule is said to be interesting or valid if its support and confidence are user specified minimum support and confidence thresholds respectively. The association rule primarily comprises two phases based on the identification of frequent item sets from the given data mining contexts. However, the problem of massive real world data transactions can be rounded off by adopting other alternatives, which in turn benefits in lossless representation of data. Theoretically, transaction database and relation database are two different intertransformable representations of data.

The production of association mining is a rule base with hundreds to millions of rules. In order to highlight the important and key ones, certain other rules are proposed which are Second -order rule which states that if the cover of the item set is known, then the corresponding relation can be easily derived. All the technical definitions given hence forth deal with the transaction of the data through item-sets in association mining. The equivalent property significantly states that rules and classes of the same hierarchical database support the power in the content. Traditional data mining techniques are implemented in order to justify the property and its corresponding definition in the specified context. The thus identified second order rules can be used to filter out useless rules out of the priority rule-set.

The effectiveness and efficiency of the classical methods in plating the rules is thoroughly verified under the 2.8 GHZ Pentium PC. Two group experiments were conducted to prune the insignificant association rules and to remove useless association classification rules. Removal of non-predictive rules by virtue of information gain metric is much similar to CHARM and CBA which also work on the same track. To generate association and classification rules by pruning method of Apriori software, some external tools are essentially required. The effectiveness of the pruning algorithm can be inversely related to the number of rules. This along with the computational time consumed, determine an efficient criterion of pruning.

Efficient post processing methods are hence proposed to remove pointless rules from rule-ways by eliminating redundancy among rules. The dependent relation, exploitation, makes this method a self manageable knowledge. The pruning procedure has been sliced into three stages starting from derivation to pruning operation on rule-set by the use of close rulesets. It is cost-effective and consumes very little time for the transaction. Hence forth, it can be applied to exploit sampling techniques and data structures, thereby increasing the efficiency.

Huawen Liu, et al [14] presented a technique on post-processing for rule reduction using closed set that was targeting to filter the otiose rules in a post-processing of rule mining. The empirical study proved that the discovery of dependent relations from closed set helps to eliminate redundant rules. Hetal Thakkar et al [12]. In the case of stream data, the post-mining of association is more challenging and continuous post mining of association rules is an unavoidable requirement, which is discussed by this author. He presented a technique for continuous post-mining of association rules in a data stream management system. He described the architecture and techniques used to achieve this advanced functionality in the Stream Mill Miner (SMM) prototype, an SQL-based DSMS designed to support continuous mining queries, which is impressive. Hacene Cherfi et al, [10] discussed a post association rule mining approach for text mining that combines data mining, semantic techniques for postmining and selection of association rules. To focus on the result analysis and to find new knowledge units, classification of association rules according to qualitative criteria using domain model as background knowledge has been introduced. The authors carried out an empirical study on molecular biology dataset that proved the benefits of taking into account a knowledge domain model of the data. Ronaldo Cristiano Prati [3]. The Receiver

Operating Characteristics (ROC) graph is a popular way of assessing the performance of classification rules, but they are inappropriate to evaluate the quality of association rules, as there is no class in association rule mining and the consequent part of different association rules might not have any correlation at all. Chapter VIII presents a novel technique of QROC, a variation of ROC space to analyze itemset costs/benefits in association rules. It can be used to help analysts to evaluate the relative interestingness among different association rules in different cost scenarios.

8 III. Attribute Relational Analysis (ARA)

Framework for coherent association rules

The approach Attribute Relational Analysis in short can refer as ARA is post mining process to prune the rules based on attribute relational relevancy. The process of ARA Framework can be classified as The process steps involved in ARA framework are 1. Initially ARA measures the property support degree for each attribute involved in given rule. 2. By using the property support degree of the attributes, Attribute Relation support of attribute pairs of the given rule will be measured. 3. With the help of Attribute Relation supports of all attribute pairs of an itemset that belongs to a given rule, Attribute relation support degree of that itemset will be measured. 4. Using these Attribute relation support degrees of Left Hand Side and Right Hand Side itemsets of the given rule, relation confidence of the rule will be determined. 5. Prunes the rules based on their attribute relation support degree. Detailed explanation of each step can be found in Section V.

IV. Closed Itemset Mining using PEPP [34] and Inference relations [35] a) Dataset adoption and formulation Item Sets I : A set of diverse elements by which the sequences generate.

9 ?

Represents a sequence's' of items those belongs to set of distinct items 'I'. 'm': total ordered items. P(e i): a transaction, where e i usage is true for that transaction.1 t S s j j = =

? S: represents set of sequences 't': represents total number of sequences and its value is volatile s j : is a sequence that belongs to S Subsequence : a sequence p s of sequence set 'S' is considered as subsequence of another sequence q s of Sequence Set 'S' if all items in sequence S p is belongs to s q as an ordered list. This can be formulated asIf ( ) ( ) 1 n s s s s pi q p q i ? ? ? = ? Then : 1 1 n m s s pi qj i j < = = ? ?

where s S and s S p q ? ? Total Support 'ts' : occurrence count of a sequence as an ordered list in all sequences in sequence set 'S' can adopt as total support 'ts' of that sequence. Total support 'ts' of a sequence can determine by fallowing formulation.

(As a first stage of the proposal we perform dataset pre-processing and itemsets Database initialization. We find itemsets with single element, in parallel prunes itemsets with single element those contains total support less than required support.

10 Forward Edge Projection

In this phase, we select all itemsets from given itemset database as input in parallel. Then we start projecting edges from each selected itemset to all possible elements. The first iteration includes the pruning process in parallel, from second iteration onwards this pruning is not required, which we claimed as an efficient process compared to other similar techniques like BIDE. In first iteration, we project an itemset p s that spawned from selected itemset i s from The above process continues till the elements available in memory those are connected through direct or transitive edges and projecting itemsets i.e., till graph become empty c) Inference Analysis [35] Inferences:

Pattern positive score is sum of no of transactions in which all items in the pattern exist, no of transactions in which all items in the pattern does not exist Pattern negative score is no of transactions in which only few items of the pattern exist Pattern actual coverage is pattern positive score-pattern negative score

11 Interest gain: Actual coverage of the pattern involved in association rule

Coherent rule Actual coverage of the rule's left side pattern must be greater than or equal to actual coverage of the right side pattern Inference Support ia s : refers actual coverage of the pattern Set I = {i1, i2? im} be the universe of items composed of m different attributes, ik(k=1,2,...,m) is item. Transaction database D is a collection of transaction T, A transaction t = (tid, X) is a tuple where tid is a unique transaction ID and X is an itemset. The count of an itemset X in D, denoted by count(X), is the number of transactions in D containing X. The support of an itemset X in D, denoted by supp(X), is the proportion of transactions in D that contain X. The negative rule X?¬Y holds in the transaction set D with confidenceconf (X?¬Y) = supp ( ) X Y ? ¬ /supp (X).

In Transaction database, each transaction is a collection of items involved sequences. The issue of mining association rules is to get all association rules that its support and confidence is respectively greater than the minimum threshold given by the user.

The issues of mining association rules can be divide into two sub-issues as follows:

? Find frequent itemsets, Generate all itemsets that support is greater than the minimum support;

? Generate association rules from frequent itemsets.

In logical analysis, the direct calculation of support logical analysis is not convenient, To calculate the support and confidence of negative associations using the support and confidence of positive association that is known: set A, B?I?A?B=??then: sup( ) 1 sup( ); sup( ) sup( ) ( );

sup( ) sup( ) sup( ) sup( ) 1 sup( ) sup( ) sup( );A A A B A Sup A B A B B A B A B A B A B ¬ = ? ? ¬ = ? ? ¬ ? = ? ? ¬ ?¬ = ? ? + ?

Based on the above formulas we perform the logical analysis to derive the actual support of the patterns that improves the rule coherency.

12 Inference analysis by example:

Let A, B?I where I is itemset generated with the association of A,B are individual items or subsets.

Under logical analysis we determine ( ) Finally we determine ( ) ts f A B ¬ ? ¬ , ( ) ts f A B ? ¬ and ( ) ts f A B ¬ ? .ia f I = ( ) ts f I + ( ) ts f A B ¬ ? ¬ - ( ) ts f A B ? ¬ -( ) ts f A B ¬ ? ;

13 Attribute Relation Analysis Framework

The proposed post mining process Attribute Relation Analysis described in detailed here. An xml based attribute class descriptor will be prepared. Fig1 shows an example descriptor. Notations description equations can found in table 1 that fallows. <class-descriptor> <properties> <property id=1> <name>prop1</name> </property> <property id=2> <name>prop2</name> </property> <property id=3> <name>prop3</name> </property> </properties> <attributes> <attribute name="item-1" id="1" properties="{list of property ids}" /> <attribute name="item-2" id="2" properties="{list of property ids}" /> <attribute name="item-3" id="3" properties="{list of property ids}" /> . . <attribute name="item-n" id="n" properties="{list of property ids}" /> </attributes> <classes> <class name="class-1" id="1" properties="{list of property ids}" /> <class name="class-2" id="2" properties="{list of property ids}" /> <class name="class-3" id="3" properties="{list of property ids}" /> . . <class name="class-n" id="n" properties="{list of property ids}" /> </classes> <child-classes> <!–parent value must be unique –> <child-class parent="class-id" child="{list of class ids}"/> . . <child-class parent="class-id" child="{list of class ids}"/> </child-classes> <relations> <!–lhs value must be unique –> <!-classes that related to a child-class also related to it's parent class –> <relation lhs="class-id" rhs="{list of class ids}" /> . . <relation lhs="class-id" rhs="{list of class ids}" /> </relations> </class-descriptor> n : is total number of attributes in the given itemset 19 pc r lhs

Pair-count of itemset, which is lhs of rule r .

14 pc r rhs

Pair-count of itemset, which is rhs of rule r . This segment focuses mainly on providing evidence on asserting the claimed assumptions that 1) The post mining framework ARA is competent enough to momentously surpass results when evaluated against other post mining techniques **??**14, 10, 12. 2) Utilization of memory and computational complexity is less when compared to other post mining techniques. 3) There is the involvement of an enhanced occurrence and a probability reduction in the memory exploitation rate with the aid of the trait equivalent prognosis and also rim snipping of the PEPP with inference analysis and ARA. This is on the basis of the surveillance done which concludes that ARA implementation is far more noteworthy and important in contrast with the likes of other notable models [10,12,14].

JAVA 1.6_20th build was employed for accomplishment of the ARA along with PEPP under inference analysis. A workstation equipped with core2duo processor, 2GB RAM and Windows XP installation was made use of for investigation of the algorithms. The parallel replica was deployed to attain the thread concept in JAVA.

Dataset Characteristics [34,35]:

We used same experiments platform described in our earlier work [34,35].Hence the dataset that we opted is Pi and it's characteristics as described in our earlier work [34,35]:

Pi is supposedly found to be a very opaque dataset, which assists in excavating enormous quantity of recurring clogged series with a profitably high threshold somewhere close to 90%. It also has a distinct element of being enclosed with 190 protein series and 21 divergent objects. Reviewing of serviceable legacy's consistency has been made use of by this dataset. Fig. 3 portrays an image depicting dataset series extent status.

In assessment with all the other regularly quoted forms like [14,12,10], Post-Processing for Rule Reduction Using Closed Set [14] has made its mark as a most preferable, superior and sealed example of post mining copy, taking in view the detailed study of the factors mainly, experts involvement, memory consumption and runtime. [34,35] Fig: **??** : A comparison report for memory usage [34, 35] Fig 5 **??** Sequence length and number of sequences at different thresholds in Pi dataset [34,35] In contrast to ARA and RRUC [14], a very intense dataset Pi is used which has petite recurrent closed series whose end to end distance is less than 10, even in the instance of high support amounting to around 90%. The diagrammatic representation displayed in Fig 3 **??**xplains that the above mentioned two algorithms execute in a similar fashion in case of rules that are generated at support 90% and above. But in situations when the support case is 88% and less, then the act of ARA surpasses RRUC

[14]'s routine. The disparity in memory exploitation of ARA and RRUC [14] can be clearly observed because of the consumption level of ARA being low than that of RRUC.

15 Conclusion

We proposed a post mining process called Attribute relation analysis framework (ARA) for pruning association rules that are contextually irrelevant. In earlier works [10,12,14] the contextual irrelevancy was identified in various ways such as (1) rule evaluation by domain expert, (2) rule evaluation by itemset closeness. We argued that none of these two models is significant in all contexts. In second case adaptability to various data contexts is missing. In the first case, rule selection highly influenced by the experts view, that is when expert changes then rule significance might be rated differently. To defend these limits here we proposed a post mining process as an extension to our earlier proposed closed itemset mining algorithm PEPP with inference analysis [34,35]. Here in this proposed post mining process ARA, the experts view is not defending [1] [2] [3]

Figure 1:

[2] , here k p is k th pair of pair-set ps of itemset rhs of rule r28 | | 1 | | PS r

[3] Global I) 2011 December (

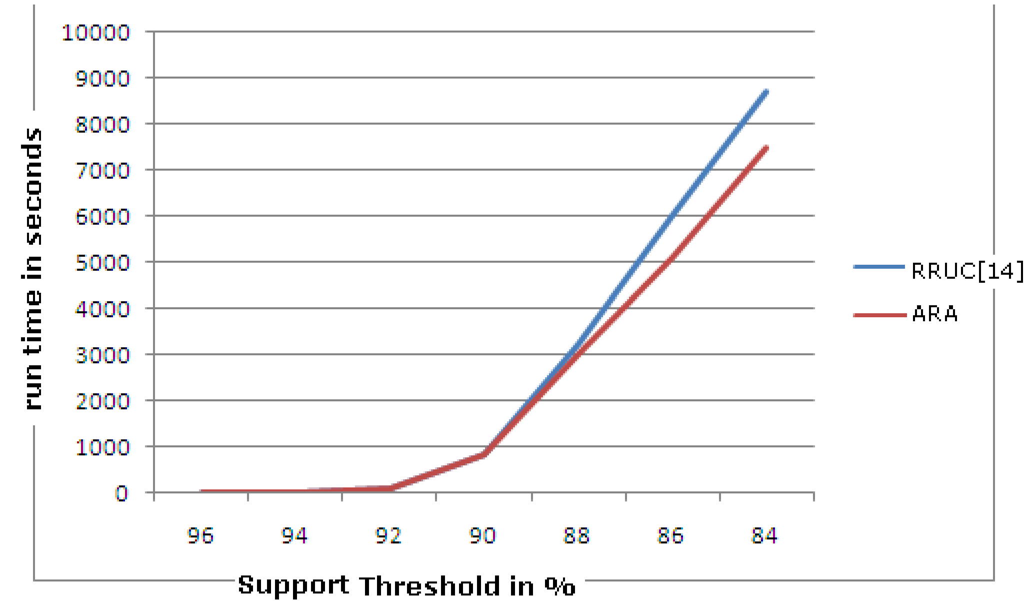


Figure 2:

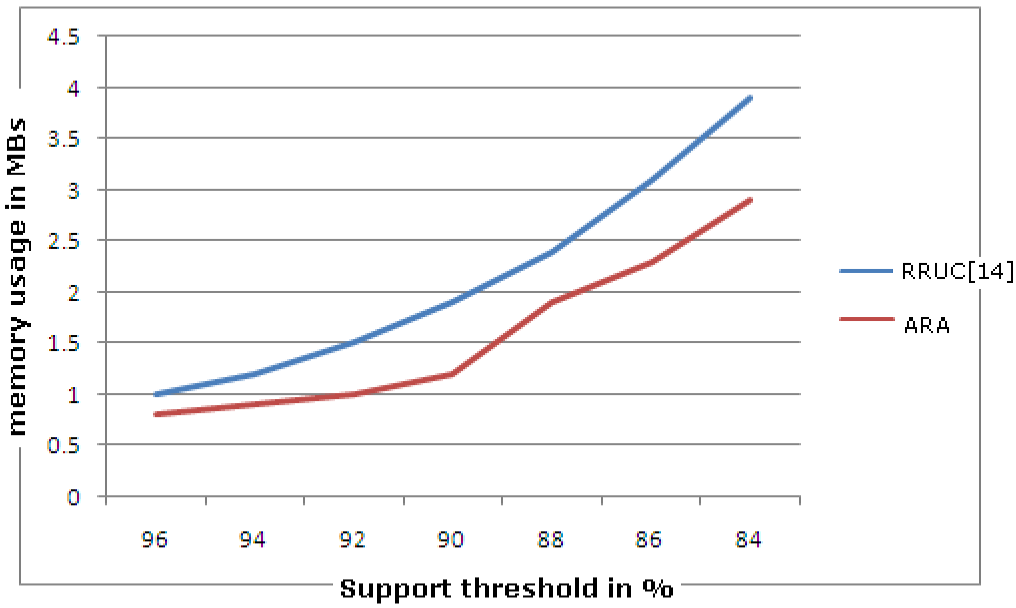


Figure 3:

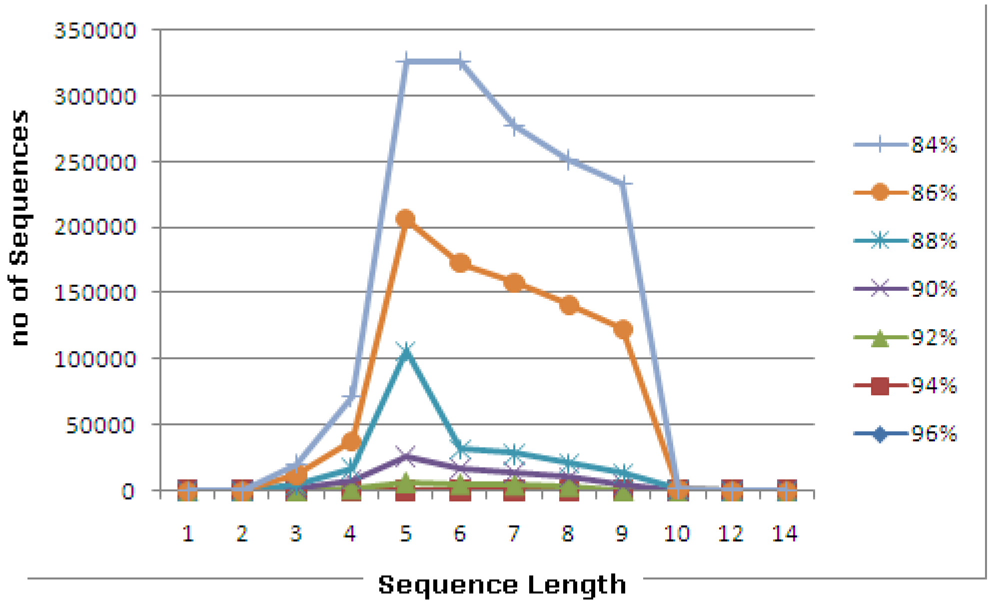
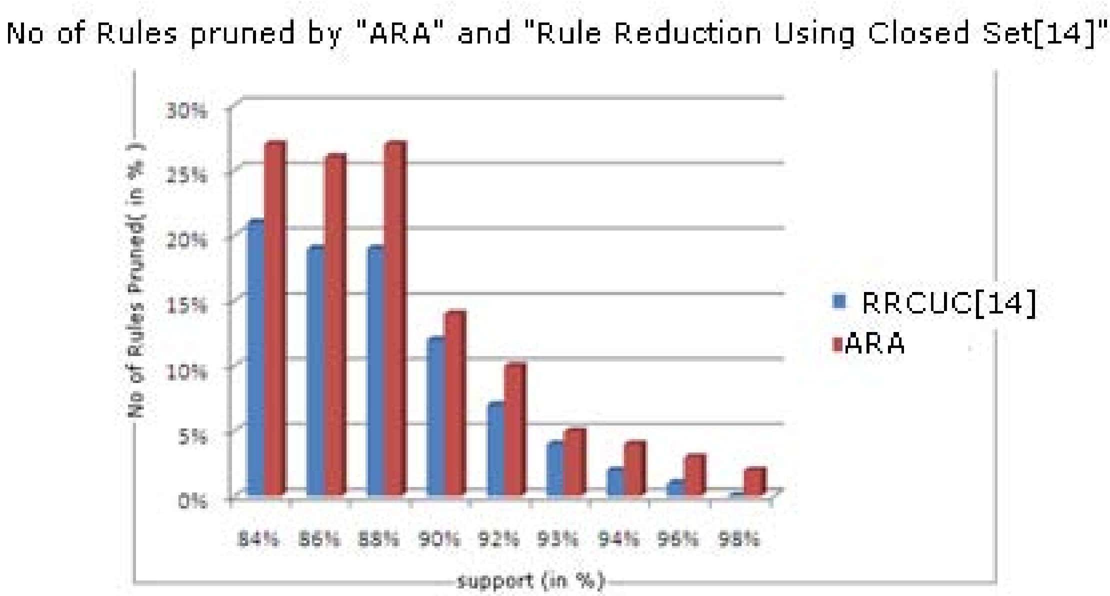


Figure 4:



3

Figure 5: 3 .

2011
December
-
I)
(

[Note: denotes which attributes should occur in the antecedent or the consequent part of the rule.]

Figure 6:

2011
December
I)
(
Global ? Closed itemset mining ? Describing item class descriptor
The input for ARA Framework is
1. A set of rules
2. An XML descriptor describes attributes, classes, class properties and class relations.
Here in this proposal we considered our earlier work [] to find closed itemsets.

Figure 7:

1

represents the notations used in ARA framework.
Class Descriptor: The domain expert classifies the attributes involved in transactions will be classified in to different categories. The process of classification as follow
1. Initially classes will be derived based on the properties; hence each class contains set of properties. These classes can be recursive i.e., a class may refer one or more other classes as sub classes.
2. Based on attribute properties, attributes will be categorized into a class.
a) Ex: if most of the attribute 'a' matched to class 'c' then a c ?

Figure 8: Table 1

1

itemset and right hand side itemset of a given rule[see table 1 row: 29] row: 8].

Property Support degree : indicates the ratio of properties matched to class level properties [table 1 row: 9].

Ex: psd a = a c ps ; cpc here a is an attribute of class c

[a c ?]

Attribute Relation support : Indicates the strength of the relation between two attributes of an itemset that are considered as pair for equation [see table 1 row: 11, 13].

Pair Count : Total number of two attributes sets; here these attribute sets must be unique [see table 1, row 17, 18]

Attribute Relation support degree : is an itemset level measurement representing average relation strength of the attributes those belongs to an itemset [see table 1 row: 24]

Relation confidence : is a rule level measurement concludes the relation strength between left hand side

[Note: [Table 1; row: 13, 15, 15 and 16] Find Attribute relation support degree ARSD r lhs of r lhs [Table 1; row: 24, 26].]

Figure 9: Table 1

1

Find Relation confidence r rc of rule r
If r rc rct ? then add rule r to resultant rule set
End
End
Fig 2 : Attribute Relation Analysis algorithm

[Note: 'RS]

Figure 10: Table 1 ;

.1 Global

[International Journal of Computer Science and Information Security (2011)] , *International Journal of Computer Science and Information Security* September 2011. September 2011. 9 (9) .

[Cherfi et al. ()] 'A Conformity Measure Using Background Knowledge for Association Rules: Application to Text Mining'. Hacene Cherfi , Amedeo Napoli , Yannick Toussaint . *Post-Mining of Association Rules: Techniques for Effective Knowledge Extraction*, 2009. p. .

[Kuntz et al. ()] 'A User-Driven Process for Mining Association Rules'. P Kuntz , F Guillet , R Lehn , H Briand . *4th Eur. Conf. on Principles of Data Mining and Knowledge Discovery*, D Zighed, H Komorowski, & J Zytkow (ed.) 2000.

[Carmona-Saez et al. ()] 'A: :Integrated analysis of gene expression by association rules discovery'. P Carmona-Saez , M Chagoyen , A Rodriguez , O Trelles , J M Carazo , Montano . *BMC Bioinformatics* 2006.

[Piatetsky-Shapiro et al. ()] 'Analysis, and Presentation of Strong Rules'. G Piatetsky-Shapiro , G Piatetsky-Shapiro , W J Frawley , S Brin , R Motwani , J D Ullman . *Proceedings of the 11th international symposium on Applied Stochastic Models and Data Analysis ASMDA*, (the 11th international symposium on Applied Stochastic Models and Data Analysis ASMDA) 2005. 16 p. .

[Mcnicholas ()] 'Association rule analysis of CAO data (with discussion)'. P Mcnicholas . *Journal of the Statistical and Social Inquiry Society of Ireland* 2007.

[Zaki and Hsiao ()] 'Charm: An Efficient Algorithm for Closed Itemset Mining'. M J Zaki , C J Hsiao . *Proc. Second SIAM Int'l Conf. Data Mining*, (Second SIAM Int'l Conf. Data Mining) 2002. p. .

[Pei et al. ()] 'Closet: An Efficient Algorithm for Mining Frequent Closed Itemsets'. J Pei , J Han , R Mao . *Proc. ACM SIGMOD Workshop Research Issues in Data Mining and Knowledge Discovery*, (ACM SIGMOD Workshop Research Issues in Data Mining and Knowledge Discovery) 2000. p. .

[Lent et al. ()] 'Clustering Association Rules'. B Lent , A N Swami , J Widom . *Proceedings of the Thirteenth International Conference on Data Engineering*, (the Thirteenth International Conference on Data EngineeringBirmingham U.K.) 1997. IEEE Computer Society.

[Thakkar et al. ()] 'Continuous Post-Mining of Association Rules in a Data Stream Management System'. Hetal Thakkar , Barzan Mozafari , Carlo Zaniolo . *Post-Mining of Association Rules: Techniques for Effective Knowledge Extraction*, 2009. p. .

[Basu et al. ()] 'Evaluating the Novelty of Text-Mined Rules using Lexical Knowledge'. S Basu , R J Mooney , K V Pasupuleti , J Ghosh . *7th ACM SIGKDD International Conference on Knowledge Discovery in Databases*, 2001.

[Kawano and Kawahara ()] 'Extended Association Algorithm Based on ROC Analysis for Visual Information Navigator'. H Kawano , M Kawahara . Lecture Notes In Computer Science 2002, 2281. p. .

[Jaroszewicz and Scheffer ()] 'Fast Discovery of Unexpected Patterns in Data, Relative to a Bayesian Network'. S Jaroszewicz , T Scheffer . *ACM SIGKDD Conference on Knowledge Discovery in Databases*, 2005.

[Klemettinen et al. ()] 'Finding Interesting Rules from Large Sets of Discovered Association Rules'. M Klemettinen , H Mannila , P Ronkainen , H Toivonen , A I Verkamo . *Proc. Int'l Conf. Information and Knowledge Management (CIKM)*, (Int'l Conf. Information and Knowledge Management (CIKM)) 1994. p. .

[Jaroszewicz and Simovici ()] 'Interestingness of Frequent Itemsets using Bayesian networks as Background Knowledge'. S Jaroszewicz , D Simovici . *ACM SIGKDD Conference on Knowledge Discovery in Databases*, 2004.

[Faure et al. ()] 'Iterative Bayesian Network Implementation by using Annotated Association Rules'. C Faure , D Delprat , J F Boulicaut , A Mille . *15th Int'l Conf. on Knowledge Engineering and Knowledge Management -Managing Knowledge in a World of Networks*, 2006.

[Marinica and Guillet ()] 'Knowledge-Based Interactive Postmining of Association Rules Using Ontologies'. Claudia Marinica , Fabrice Guillet . *IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING* JUNE 2010. 22 (6) .

[Machine Learning, and Knowledge Discovery in Databases ()] *Machine Learning, and Knowledge Discovery in Databases*, 1995. p. . (Workshop Statistics)

[Burdick et al. (2005)] 'Mafia: A Maximal Frequent Itemset Algorithm'. D Burdick , M Calimlim , J Flannick , J Gehrke , T Yiu . *IEEE Trans. Knowledge and Data Eng* Nov. 2005. 17 (11) p. .

[Mining Closed Itemsets for Coherent Rules: An Inference Analysis Approach By Kalli Srinivasa Nageswara Prasad Double Blind 'Mining Closed Itemsets for Coherent Rules: An Inference Analysis Approach By Kalli Srinivasa Nageswara Prasad'. *Double Blind Peer Reviewed International Research Journal Publisher: Global Journals Inc. (USA) Online* 0975-4172 and Print : 0975. 19 Version 1.0 November 2011. 11. (Prof. S. Ramakrishna, Global Journal of Computer Science and Technology)

[Savasere et al. ()] 'Mining for strong negative associations in a large database of customer transactions'. A Savasere , E Omiecinski , S Navathe . *Proceedings of the 14th International Conference on Data Engineering*, (the 14th International Conference on Data EngineeringWashington DC, USA) 1998.

[Srikant ()] 'Mining Generalized Association Rules'. R Srikant , R . http://citeseer.ist.psu.edu/srikant95mining.html *Proc. 21st Int'l Conf. Very Large Databases*, (21st Int'l Conf. Very Large Databases) 1995. p. .

[Bayardo and Roberto ()] 'Mining the Most Interesting Rules'. J Bayardo , J Roberto , R . *Proc. ACM SIGKDD*, (ACM SIGKDD) 1999. p. .

[Li (2006)] 'On Optimal Rule Discovery'. J Li . *IEEE Trans. Knowledge and Data Eng* Apr. 2006. 18 (4) p. .

[Parallel Edge Projection and Pruning (PEPP) Based Sequence Graph protrude approach for Closed Itemset Mining kalli Sriniva *Parallel Edge Projection and Pruning (PEPP) Based Sequence Graph protrude approach for Closed Itemset Mining kalli Srinivasa Nageswara Prasad*, ol. 9. Tirupati, Andhra Pradesh , India; Tirupati, Andhra Pradesh , India. Sri Venkateswara University ; Department of Computer Science, Sri Venkateswara University (Prof. S. Ramakrishna)

[Liu and Hsu ()] *Post-Analysis of Learned Rules*, B Liu , W Hsu . 1996. AAAI/IAAI.

[Liu et al. ()] 'Post-Processing for Rule Reduction Using Closed Set'. Huawen Liu , Jigui Sun , Huijie Zhang . *Post-Mining of Association Rules: Techniques for Effective Knowledge Extraction*, 2009. p. .

[Fawcett ()] 'PRIE: A system to generate rulelists to maximize ROC performance'. T Fawcett . *Data Mining and Knowledge Discovery Journal* 2008. 17 (2) p. .

[Toivonen et al.] 'Pruning and Grouping of Discovered Association Rules'. H Toivonen , M Klemettinen , P Ronkainen , K Hatonen , H Mannila . *Proc. ECML-95*, (ECML-95)

[Liu et al.] 'Pruning and summarizing the discovered associations'. B Liu , W Hsu , Y Ma . *Proceedings of the fifth ACM SIGKDD international conference on Knowledge discovery and data mining 1999*, (the fifth ACM SIGKDD international conference on Knowledge discovery and data mining 1999) p. .

[Liu et al. ()] 'Pruning and summarizing the discovered associations'. B Liu , W Hsu , Y Ma . *Proceedings of the fifth ACM SIGKDD international conference on Knowledge discovery and data mining*, (the fifth ACM SIGKDD international conference on Knowledge discovery and data mining) 1999.

[Ronaldo Cristiano ()] 'QROC: A Variation of ROC Space to Analyze Item Set Costs/Benefits in Association Rules'. Prati Ronaldo Cristiano . *Post-Mining of Association Rules: Techniques for Effective Knowledge Extraction*, 2009. p. .

[Prati and Flach ()] 'ROCCER: an algorithm for rule learning based on ROC analysis'. R Prati , P Flach . *Proceeding of the 19th International Joint Conference on Artificial Intelligence*, (eeding of the 19th International Joint Conference on Artificial Intelligence) 2005.

[Hahsler et al. ()] 'Selective Association Rule Generation'. M Hahsler , C Buchta , K Hornik . *Computational Statistic* 2008. Kluwer Academic Publishers. 23 (2) p. .

[Fienberg and Shmueli ()] 'Statistical issues and challenges associated with rapid detection of bioterrorist attacks'. S E Fienberg , G Shmueli . *Statistics in Medicine* 2005.

[Zaki and Ogihara (1998)] 'Theoretical Foundations of Association Rules'. M J Zaki , M Ogihara . *Proc. Workshop Research Issues in Data Mining and Knowledge Discovery (DMKD '98)*, (Workshop Research Issues in Data Mining and Knowledge Discovery (DMKD '98)) June 1998. p. .

[Mozafari et al. (2008)] 'Verifying and Mining Frequent Patterns from Large Windows over Data Streams'. Barzan Mozafari , Hetal Thakkar , Carlo Zaniolo . *Proceedings of the 24th International Conference on Data Engineering*, (the 24th International Conference on Data EngineeringCancún, México) 2008. April 7-12, 2008.