M.Saritha, T.Venkata Ramana, Kareemunnisa, Y.Sowjanya, M.Nishantha / International Journal of Engineering Research and Applications (IJERA) ISSN: 2248-9622 www.ijera.com
Vol. 2, Issue 5, September- October 2012, pp.1342-1351

# Attribute Ontological Relational Weights (AORW) For Coherent Association Rules: A Post Mining Process For Association Rules

**M.Saritha[1], T.Venkata Ramana[2], Kareemunnisa[3], Y.Sowjanya[4], M.Nishantha[5]**

[1](Asst. Professor, Department of Computer Science, GuruNanak Institutions of Technical Campus, India)
[2](Professor, Head of the Department of Computer Science, SLC's IET, India)
[3, 5] (Asst. Professors, Department of Information Technology, GuruNanak Institutions of Technical Campus, India)
[4] (M.tech student, Department of Computer Science and Engg, SLC's IET, India)

## 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 AORW as an extension to closed itemset mining algorithms.**

**Keywords**: post mining, association rule mining, closed itemset, PEPP, Inference analysis, rule pruning.

## 1. INTRODUCTION

In general, association rules tend to deliver an efficient method of analysing binary or discredited 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 $X \Rightarrow Y$, where $X, Y \subset I$ such that $X \neq \emptyset$, $Y \neq \emptyset$, and $X \cap Y = \emptyset$. 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. X 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} $\Rightarrow$ {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 $X \not\Rightarrow Y$ [20]. Another is to think in terms of negations. Suppose $X \subset I$, then write $\neg X$ to denote the absence or negation of the item, or items, in X from a transaction. Considering X as a binary {0, 1} variable, the presence 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 [20].

**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 [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 [19] proposed clustering the discovered association rules.

**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 expert's 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.

## 2. RELATED WORK

Huawen Liu, [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, aprori knowledge has also been taken into account in literatures and a template denotes which attributes should occur in the antecedent or the consequent part of the rule.

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 inter-transformable 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 rule-sets. 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, [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 [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 [10] discussed a post association rule mining approach for text mining that combines data mining, semantic techniques for post-mining 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.

## 3. ATTRIBUTE ONTOLOGICAL RELATIONAL WEIGHTS FOR COHERENT ASSOCIATION RULES

The approach Attribute Ontological Relational Weights in short can refer as AORW is post mining process to prune the rules based on attribute relational relevancy. The process of AORW Framework can be classified as

- Closed itemset mining
- Describing item class descriptor

The input for AORW 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.

The process steps involved in AORW framework are

1. Initially AORW 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.

## 4. CLOSED ITEMSET MINING [34]

### A. Dataset adoption and formulation

Item Sets I: A set of diverse elements by which the sequences generate.

$$I = \bigcup_{k=1}^{n} i_k$$

Note: 'I' is set of diverse elements

Sequence set 'S': A set of sequences, where each sequence contains elements each element 'e' belongs to 'I' and true for a function p(e). Sequence set can formulate as

$$s = \bigcup_{i=1}^{m} <e_i | (p(e_i), e_i \in I)>$$

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.

$$S = \bigcup_{j=1}^{t} s_j$$

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 $s_p$ of sequence set 'S' is considered as subsequence of another sequence $s_q$ of Sequence Set 'S' if all items in sequence $S_p$ is belongs to $s_q$ as an ordered list. This can be formulated as

If $$(\bigcup_{i=1}^{n} s_{pi} \in s_q) \Rightarrow (s_p \subseteq s_q)$$

Then $$\bigcup_{i=1}^{n} s_{pi} <: \bigcup_{j=1}^{m} s_{qj} \qquad s_p \in S \text{ and } s_q \in S$$ where

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.

$$f_{ts}(s_t) = |s_t <: s_p \ (\text{for each } p = 1..|DB_S|)|$$

$DB_S$ Is set of sequences

$f_{ts}(s_t)$ : Represents the total support 'ts' of sequence $s_t$ is the number of super sequences of $s_t$

Qualified support '$q_s$': The resultant coefficient of total support divides by size of sequence database adopt as qualified support 'qs'. Qualified support can be found by using fallowing formulation.

$$f_{qs}(s_t) = \frac{f_{ts}(s_t)}{|DB_S|}$$

Sub-sequence and Super-sequence: A sequence is sub sequence for its next projected sequence if both sequences having same total support.

Super-sequence: A sequence is a super sequence for a sequence from which that projected, if both having same total support.

Sub-sequence and super-sequence can be formulated as

If $f_{ts}(s_t) \geq rs$ where 'rs' is required support threshold given by user

And $s_t <: s_p$ *for any p value* where $f_{ts}(s_t) \cong f_{ts}(s_p)$

### B. Closed Itemset Discovery[34]

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.

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 $s_p$ that spawned from selected itemset $s_i$ from $DB_S$ and an element $e_i$ considered from 'I'. If the $f_{ts}(s_p)$ is greater or equal to $rs$, then an edge will be defined between $s_i$ and $e_i$. If $f_{ts}(s_i) \cong f_{ts}(s_p)$ then we prune $s_i$ from $DB_S$. This pruning process required and limited to first iteration only.

From second iteration onwards project the itemset $S_p$ that spawned from $S_{p'}$ to each element $e_i$ of 'I'. An edge can be defined between $S_{p'}$ and $e_i$ if $f_{ts}(s_p)$ is greater or equal to $rs$. In this description $S_{p'}$ is a projected itemset in previous iteration and eligible as a sequence. Then apply the fallowing validation to find closed sequence.

Edge pruning :

If any of $f_{ts}(s_{p'}) \cong f_{ts}(s_p)$ that edge will be pruned and all disjoint graphs except $s_p$ will be considered as closed sequence and moves it into $DB_S$ and remove all disjoint graphs from memory.

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

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

$f_{ia}(s_t)$ : Represents the inference support of the sequence $s_t$

Approach:

For each pattern $s_p$ of the pattern dataset, If $f_{ia}(s_t) < ia_s$ then we prune that pattern

## 5. ATTRIBUTE ONTOLOGICAL RELATIONAL WEIGHTS FRAMEWORK

The proposed post mining process Attribute Ontological Relational Weights described in detailed here. Table 1 represents the notations used in AORW framework.

.

Table1: Notations used in Attribute Ontological Relational Weights

| | | |
|---|---|---|
| 1 | $r_{lhs}$ | Left side itemset of the Rule $r$ |
| 2 | $r_{rhs}$ | Right side itemset of the rule $r$ |
| 3 | $RS$ | Rule set |
| 4 | $cpc_c$ | Class properties count of class $c$ |
| 5 | $apc_a$ | Attribute property count of attribute $a$ |
| 6 | $psd$ | Property support degree |
| 7 | $tp_c$ | Total properties of class c |
| 8 | $ps_a$<br>$ps_a = \frac{apc_a}{cpc_c}$ | Property support of attribute a of class $c$ |
| 9 | $psd_a = \frac{ps_a}{tp_c}$ | where $a \in c$ |
| 10 | $rs_c$ | Relation support of class $c$ is max threshold value 1. |
| 11 | $ARS$ | Attribute Relation support |
| 12 | $ARSD_i$ | Attribute Relation support Degree of itemset $i$. |
| 13 | If attributes $a_i$ and $a_j$ belongs to same class $c$ | $ARS(a_i, a_j) \cong 1$, $where\ \{a_i, a_j\} \in c$ |
| 14 | If attribute $a_i$ belongs to class $c_i$ and attribute $a_j$ belongs to class $c_j$, $c_i$ and $c_j$ relation is true then | $ARS(a_i, a_j) = \frac{psd_{a_i} + psd_{a_j}}{rs_{c_i} + rs_{c_j}}$ |
| 15 | If $c_i$ and $c_j$ relation is false | $ARS(a_i, a_j) = 0$, |
| 16 | $ARS(a_i, a_j) \cong ARS(a_j, a_i)$ | Applicable in all cases such as both belongs to same class, both belongs to different classes that are not having relation and both belongs to different classes that are having relation |
| 17 | No of attribute pairs in an itemset | $pc$ |

| | | |
|---|---|---|
| 18 | $pc = 0;$ <br> $\sum_{i=1}^{n-1} pc + i \quad (or) \quad pc = \frac{n-1}{2} \times n$ | $pc$ : is total number of pairs in a given itemset. <br> $n$ : is total number of attributes in the given itemset |
| 19 | $pc_{r_{lhs}}$ | Pair-count of itemset, which is $lhs$ of rule $r$ . |
| 20 | $pc_{r_{rhs}}$ | Pair-count of itemset, which is $rhs$ of rule $r$ . |
| 21 | $pc_{r_{lhs} \cup r_{rhs}}$ | Pair-count of itemset that generated from $r_{lhs} \cup r_{rhs}$ |
| 22 | $PS_i = \{p_1, p_2, ......., p_m\}$ | Pair set that generated from itemset $i$ |
| 23 | $ARS(p_i)$ | Attribute relation support of pair $p_i$ . |
| 24 | $ARSD_i = \frac{\sum_{k=1}^{\lvert ps_i \rvert} ARS(p_k)}{\lvert ps_i \rvert}$ | Attribute relation support degree of itemset $i$ <br> And $p_k \in ps_i$ , here $p_k$ is k[th] pair of pair-set $ps$ of itemset $i$ |
| 25 | $RC_r$ | Relation confidence of rule $r$ . |
| 26 | $ARSD_{r_{lhs}} = \frac{\sum_{k=1}^{\lvert PS_{r_{lhs}} \rvert} p_k}{\lvert PS_{r_{lhs}} \rvert}$ | Attribute relation support degree of itemset, which is $lhs$ of rule $r$ <br> And $p_k \in ps_{r_{lhs}}$ , here $p_k$ is k[th] pair of pair-set $ps$ of itemset $lhs$ of rule $r$ |
| 27 | $ARSD_{r_{rhs}} = \frac{\sum_{k=1}^{\lvert PS_{r_{rhs}} \rvert} p_k}{\lvert PS_{r_{rhs}} \rvert}$ | Attribute relation support degree of itemset, which is $rhs$ of rule $r$ <br> And $p_k \in ps_{r_{rhs}}$ , here $p_k$ is k[th] pair of pair-set $ps$ of itemset $rhs$ of rule $r$ |
| 28 | $ARSD_{r_{lhs} \cup r_{rhs}} = \frac{\sum_{k=1}^{\lvert PS_{r_{lhs} \cup r_{rhs}} \rvert} p_k}{\lvert PS_{r_{lhs} \cup r_{rhs}} \rvert}$ | Attribute relation support degree of itemset that generated from $lhs \bigcup rhs$ of rule $r$ <br> And $p_k \in PS_{r_{lhs} \cup r_{rhs}}$ , here $p_k$ is k[th] pair of pair-set $ps$ of itemset that generated from $lhs \bigcup rhs$ of rule $r$ |
| 29 | $rc_r = \frac{ARSD_{r_{lhs} \cup r_{rhs}}}{ARSD_{r_{lhs}}}$ | Relation confidence of $r$ is the coefficient emerged as result when 30attribute support degree of all attribute involved in rule $r$ is divided by attribute support degree of rule $r$ 's $lhs$ |

**Property Support:** No of attribute properties are matched to number class properties to which that attribute belongs to [Table 1 row: 8].

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

Ex: $psd_a = \frac{ps_a}{cpc_c}$ ; here $a$ is an attribute of class $c\ [a \in c]$

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

**AORW algorithm:**

Input: Rule set $RS$ , Class Descriptor $CD$ and relation confidence threshold $rct$

Output: Significant Rule set $RS'$ which is subset of $RS$ $RS' \subseteq RS$

Begin:

While $RS$ is not empty

Begin:

Read a rule $r$ from $RS$

Find property support $ps_a$ and property support degree $psd_a$ of each attribute $a$ of $r_{lhs}$ [Table 1 row: 8, 9]

Find property support $ps_a$ and property support

degree $psd_a$ of each attribute $a$ of $r_{rhs}$ [Table 1 row: 8, 9]
Find unique two attribute pair set $PS_{\eta_{hs}}$ from $r_{lhs}$ [Table 1 row: 22]
Find Attribute relation support $ARS_p$ for each pair $p$, where $p \in PS_{r_{lhs}}$ [Table 1; row: 13, 15, 15 and 16]
Find Attribute relation support degree $ARSD_{\eta_{hs}}$ of $\eta_{hs}$ [Table 1; row: 24, 26].
Find unique two attribute pair set $PS_{\eta_{hs}}$ from $r_{lhs}$ [Table 1 row: 22]
Find Attribute relation support $ARS_p$ for each pair $p$, where $p \in PS_{r_{rhs}}$ [Table 1; row: 13, 15, 15 and 16]
Find Attribute relation support degree $ARSD_{r_{rhs}}$ of $\eta_{hs}$ [Table 1; row: 24, 27].
Find unique two attribute pair set $PS_{\eta_{hs} \cup r_{rhs}}$ from $r_{lhs} \cup r_{rhs}$ [Table 1 row: 22]
Find Attribute relation support $ARS_p$ for each pair $p$, where $p \in PS_{r_{lhs} \cup r_{rhs}}$ [Table 1; row: 13, 15, 15 and 16]
Find Attribute relation support degree $ARSD_{\eta_{hs} \cup r_{rhs}}$ of $r_{lhs} \cup r_{rhs}$ [Table 1; row: 24, 28].
Find unique two attribute pair set $PS_{\eta_{hs} \cup r_{rhs}}$ from $r_{lhs} \cup r_{rhs}$ [Table 1 row: 22]
Find Attribute relation support $ARS_p$ for each pair $p$, where $p \in PS_{r_{rhs} \cup \eta_{hs}}$ [Table 1; row: 13, 15, 15 and 16]
Find Attribute relation support degree $ARSD_{r_{rhs} \cup \eta_{hs}}$ of $r_{lhs} \cup r_{rhs}$ [Table 1; row: 24, 28].
Find Relation confidence $rc_r$ of rule $r$
If $rc_r \geq rct$ then add rule $r$ to resultant rule set $RS'$
End
End

Fig 2: Attribute Ontological Relational Weights algorithm

This segment focuses mainly on providing evidence on asserting the claimed assumptions that 1) The post mining framework AORW 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 AORW. This is on the basis of the surveillance done which concludes that AORW 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 AORW 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:**

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.

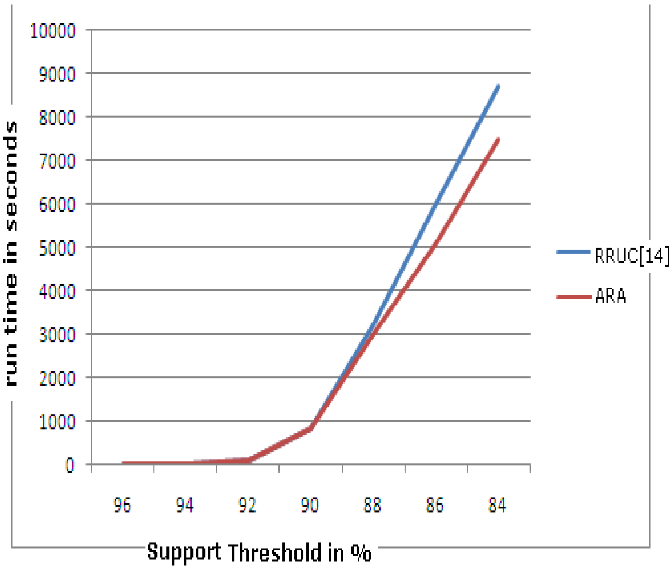


Fig 3: A comparison report for Runtime

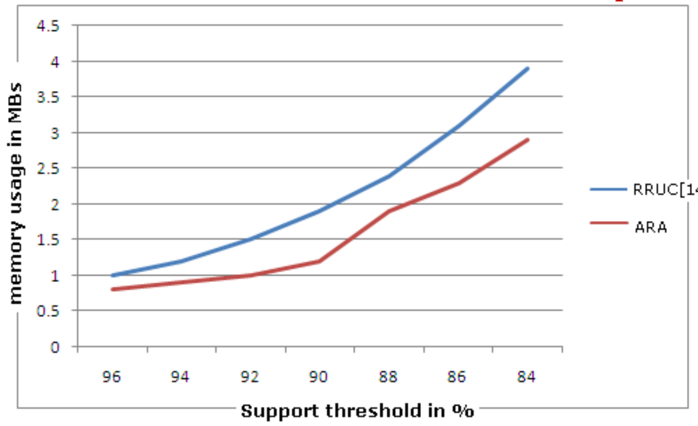


Fig: 4: A comparison report for memory usage

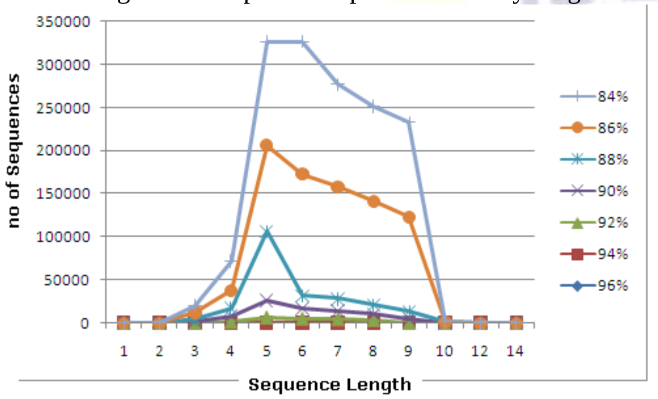


Fig 5: Sequence length and number of sequences at different thresholds in Pi dataset

In contrast to AORW 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 explains 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 AORW surpasses RRUC [14]'s routine. The disparity in memory exploitation of AORW and RRUC [14] can be clearly observed because of the consumption level of AORW being low than that of RRUC. Fig 5 indicates that rules that are contextually irrelevant have been pruned by AORW in high probable rate and stable. Apart from the benefits observed, the rules identified by AORW are more relevant to transaction consequences. It becomes possible since there is no task of experts evaluation in post mining process. Due to the concept of attribute class relation descriptor, the relation between attributes involved in rule is stable.

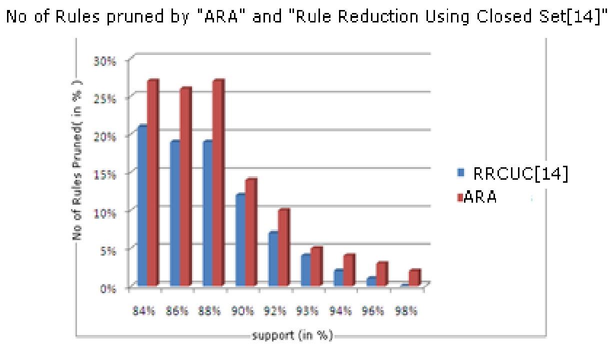


Fig 6: Rules pruned in in % by AORW and RRUC[14]

## 6. CONCLUSION

We proposed a post mining process called Attribute relation analysis framework (AORW) 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 AORW, the experts view is not defending one to other, rather it extends or refines. This become possible in AORW because of the proposed concept called attribute class relation descriptor. In this work we consider relation confidence as bench mark for rule pruning; in future this work can be extended to prune the rules by opting relation confidence threshold under inference analysis.

## REFERENCES

[1] Fienberg, S. E. & Shmueli, G: "Statistical issues and challenges associated with rapid detection of bio-terrorist attacks. Statistics in Medicine, 2005

[2] Carmona-Saez, P., Chagoyen, M., Rodriguez, A., Trelles, O., Carazo, J. M., & Pascual-Montano, A: :Integrated analysis of gene expression by association rules discovery". BMC Bioinformatics, 2006

[3] Ronaldo Cristiano Prati, "QROC: A Variation of ROC Space to Analyze Item Set Costs/Benefits in Association Rules", Post-Mining of Association Rules: Techniques for Effective Knowledge Extraction, IGIGlobal, 2009 Pages: 133-148 pp.

[4] McNicholas, P. D: "Association rule analysis of CAO data (with discussion). Journal of the Statistical and Social Inquiry Society of Ireland, 2007

[5] Prati, R., & Flach, P. "ROCCER: an algorithm for rule learning based on ROC analysis". Proceeding of the 19th International Joint Conference on Artificial Intelligence, 2005

[6] Fawcett, T: "PRIE: A system to generate rulelists to maximize ROC performance". Data Mining and Knowledge Discovery Journal, 2008, 17(2), 207-224.

[7] Kawano, H., & Kawahara, M: "Extended Association Algorithm Based on ROC Analysis for Visual Information Navigator". Lecture Notes In Computer Science, 2002, 2281, 640-649.

[8] Piatetsky-Shapiro, G., Piatetsky-Shapiro, G., Frawley, W. J., Brin, S., Motwani, R., Ullman, J. D., : "Discovery, Analysis, and Presentation of Strong Rules". Proceedings of the 11th international symposium on Applied Stochastic Models and Data Analysis ASMDA, 2005, 16, 191-200

[9] Liu, B., Hsu, W., & Ma, Y: "Pruning and summarizing the discovered associations". Proceedings of the fifth ACM SIGKDD international conference on Knowledge discovery and data mining 1999, (pp. 125-134)

[10] Hacene Cherfi, Amedeo Napoli, Yannick Toussaint: "A Conformity Measure Using Background Knowledge for Association Rules: Application to Text Mining", Post-Mining of Association Rules: Techniques for Effective Knowledge Extraction, IGIGlobal, 2009 Pages: 100-115 pp.

[11] Liu, B., Hsu, W., & Ma, Y.: "Pruning and summarizing the discovered associations". Proceedings of the fifth ACM SIGKDD international conference on Knowledge discovery and data mining, 1999

[12] Hetal Thakkar, Barzan Mozafari, Carlo Zaniolo: "Continuous Post-Mining of Association Rules in a Data Stream Management System", Post-Mining of Association Rules: Techniques for Effective Knowledge Extraction, IGIGlobal, 2009 Pages: 116-132 pp.

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

[14] Huawen Liu, Jigui Sun, Huijie Zhang: "Post-Processing for Rule Reduction Using Closed Set", Post-Mining of Association Rules: Techniques for Effective Knowledge Extraction, IGIGlobal,2009 Pages: 81-99 pp.

[15] Jaroszewicz, S., & Simovici, D. A: "Interestingness of Frequent Itemsets using Bayesian networks as Background Knowledge". ACM SIGKDD Conference on Knowledge Discovery in Databases, 2004

[16] Jaroszewicz, S., & Scheffer, T: "Fast Discovery of Unexpected Patterns in Data, Relative to a Bayesian Network". ACM SIGKDD Conference on Knowledge Discovery in Databases, 2005

[17] Faure, C., Delprat, D., Boulicaut, J. F., & Mille, A: "Iterative Bayesian Network Implementation by using Annotated Association Rules." 15th Int'l Conf. on Knowledge Engineering and Knowledge Management – Managing Knowledge in a World of Networks, 2006

[18] Liu, B., & Hsu, W: "Post-Analysis of Learned Rules". AAAI/IAAI, 1996

[19] Lent, B., Swami, A. N., & Widom, J: "Clustering Association Rules". Proceedings of the Thirteenth International Conference on Data Engineering, 1997, Birmingham U.K., IEEE Computer Society

[20] Savasere, A., Omiecinski, E., & Navathe, S. B: "Mining for strong negative associations in a large database of customer transactions". Proceedings of the 14th International Conference on Data Engineering, Washington DC, USA, 1998

[21] Basu, S., Mooney, R. J., Pasupuleti, K. V., & Ghosh J: "Evaluating the Novelty of Text-Mined Rules using Lexical Knowledge". 7th ACM SIGKDD International Conference on Knowledge Discovery in Databases, 2001

[22] Claudia Marinica and Fabrice Guillet: "Knowledge-Based Interactive Postmining of Association Rules Using Ontologies", IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 22, NO. 6, JUNE 2010

[23] M.J. Zaki and C.J. Hsiao, "Charm: An Efficient Algorithm for Closed Itemset Mining," Proc. Second SIAM Int'l Conf. Data Mining, pp. 34-43, 2002.

[24] J. Pei, J. Han, and R. Mao, "Closet: An Efficient Algorithm for Mining Frequent Closed Itemsets," Proc. ACM SIGMOD Workshop Research Issues in Data Mining and Knowledge Discovery, pp. 21-30, 2000.

[25] M.J. Zaki and M. Ogihara, "Theoretical Foundations of Association Rules," Proc. Workshop Research Issues in Data Mining and Knowledge Discovery (DMKD '98), pp. 1-8, June 1998.

[26] D. Burdick, M. Calimlim, J. Flannick, J. Gehrke, and T. Yiu, "Mafia: A Maximal Frequent Itemsgorithm," IEEE Trans. Knowledge and Data Eng., vol. 17, no. 11, pp. 1490-1504, Nov. 2005.

[27] J. Li, "On Optimal Rule Discovery," IEEE Trans. Knowledge and Data Eng., vol. 18, no. 4, pp. 460-471, Apr. 2006.

[28] M. Hahsler, C. Buchta, and K. Hornik, "Selective Association Rule Generation," Computational Statistic, vol. 23, no. 2, pp. 303-315, Kluwer Academic Publishers, 2008.

[29] H. Toivonen, M. Klemettinen, P. Ronkainen, K. Hatonen, and H. Mannila, "Pruning and Grouping of Discovered Association Rules," Proc. ECML-95 Workshop Statistics, Machine Learning, and Knowledge Discovery in Databases, pp. 47-52, 1995.

[30] J. Bayardo, J. Roberto, and R. Agrawal, "Mining the Most Interesting Rules," Proc. ACM SIGKDD, pp. 145-154, 1999.

[31] M. Klemettinen, H. Mannila, P. Ronkainen, H. Toivonen, and A.I. Verkamo, "Finding Interesting Rules from Large Sets of Discovered Association Rules," Proc. Int'l Conf. Information and Knowledge Management (CIKM), pp. 401-407, 1994.

[32] R. Srikant and R. Agrawal, "Mining Generalized Association Rules," Proc. 21st Int'l Conf. Very Large Databases, pp. 407-419, http://citeseer.ist.psu.edu/srikant95mining.html, 1995.

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

[34] Parallel Edge Projection and Pruning (PEPP) Based Sequence Graph protrude approach for Closed Itemset Mining kalli Srinivasa Nageswara Prasad, Sri Venkateswara University, Tirupati, Andhra Pradesh, India. Prof. S. Ramakrishna, Department of Computer Science, Sri Venkateswara University, Tirupati, Andhra Pradesh, India ol. 9 No. 9 September 2011 International Journal of Computer Science and Information Security Publication September 2011, Volume 9 No. 9

[35] Mining Closed Itemsets for Coherent Rules: An Inference Analysis Approach By Kalli Srinivasa Nageswara Prasad, Prof. S. Ramakrishna, Global Journal of Computer Science and Technology Volume 11 Issue 19 Version 1.0 November 2011 Type: Double Blind Peer Reviewed International Research Journal Publisher: Global Journals Inc. (USA) Online ISSN: 0975-4172 & Print ISSN: 0975-4350

## SHORT BIO DATA FOR THE AUTHOR

**Ms. M. Saritha** working as Assistant Professor in the Department of Computer Science and Engineering. Guru Nanak Institutions of Technical Campus with a teaching experience of 1.5 years. She is a B.Tech(ECE), and M.Tech in Computer science (PC) and Her areas of interest includes Mobile Computing, DataMining and Information Security.

**Professor.ThaduriVenkataRamana** received the Master Degree in Computer Science and Engineering from Osmania University, Hyderabad. Pursuing Ph.D in Computer Science and Engineering from Jawaharlal Nehru Technological University, Hyderabad (JNTUH), Hyderabad. . His area of interest in research is Information Retrieval & Data mining. Presently working as a Professor & Head, Department of Computer Science and Engineering, SLC's Institute of Engineering and Technology, Hyderabad.

**Ms. Kareemunnisa** working as Assistant Professor in the Department of Computer Science and Engineering. Guru Nanak Institutions of Technical Campus with a teaching experience of 6 years. She is a B.Tech(IT), and M.Tech in Computer scienceand Engineering (CSE) and Her areas of interest includes Mobile Computing, Data Mining and Information Security.

**Ms. Y. Sowjanya** received Bachelor Degree in Information Technology from Jawaharlal Nehru Technological University, Hyderabad (JNTUH) and pursuing Master Degree in Computer Science and Engineering (CSE) form Jawaharlal Nehru Technological University, Hyderabad (JNTUH). Her research interest includes Computer Networks and Information Retrieval Systems, Data Mining.

**Ms. M. Nishantha** working as Assistant Professor in the Department of Computer Science and Engineering. Guru Nanak Institutions of Technical Campus with a teaching experience of 6 years. She is a B.Tech(IT), and M.Tech in Computer scienceand Engineering (IT) and Her areas of interest includes Mobile Computing, DataMining and Information Security.