CHAPTER 3

RAPID SEARCH BASED ON FP-GROWTH ALGORITHM

3.1 INTRODUCTION

This chapter concentrates on the complexity of the worth of the set of association rules. It is very important because real-life databases surrender lot of the time, numerous thousands of rules with high assurance and suggest new algorithms based on closed sets to decrease the mining to bases for correct and predictable rules. On one occasion a frequent closed itemsets which is form a generating set for both frequent itemsets then the association rules have been automatically revealed.

The proposed algorithm RSDCA is efficiently generating the bases for association rules. All association rules can be revealed from the set of non-redundant rules and automatically it find all useful information. Hence, the repeated and useless rules are deleted from the set; the size of association rule is considerably reduced, when it compared with the set of all possible rules. This type of approach has an advantage on one hand that the user is provided with a smaller set of resulting rules and it is easier to handle, and also the quality of information is increased for searching the frequent itemsets. Also, the execution times of the rules are reduced compared with the other newly developed association rules.
The approach re-presented in this research belongs to the second trend because it goal is to extract all possible rules and a sub-set called foundation or coat for association rules. While computing such type of a basis, the redundant rules are discarded because they don’t have related knowledge. This type of operations is a mile-stone to the rule extraction, and considerably reduces the resulting set.

The relationships between data items or attributes are shown by the association rule and also it finds the correct relationships between them in the database. Normal accurate measures are support and confidence that spot the quantity of database transactions or objects and maintain each rule out. While the association rules support and confidence is more than some user-defined minimum thresholds, that rule is measured as relevant and that extracted knowledge could be used for supporting decision making.

To improve the efficiency of the rule discovery lot of methods were discussed. While point out the described problem, some important methods can be distinguished, which are filtering rules and Boolean operators, for selecting rules from the given items. A related approach extended with a measure of value of extracted rules, called improvement, and is planned an SQL-like operator called Mine Rule. It allows the design of general extraction criteria.
On one occasion vast amount of rules are extracted, querying facilities make it possible to handle rule subsets selected based on the user preferences known as Apriori. In contrast, to minimize the number of exhibited rules the trend addresses the problem with Apriori vision.

This work tells the basis of new method since it makes it possible by avoiding handling of large sets of rule to generate the bases from frequent closed itemsets. Apriori-Close is a frequent closed itemsets from frequent itemsets without accessing the dataset, and it extends the Apriori algorithm by discovering concurrently frequent itemsets and frequent closed itemsets without any additional execution time.

The frequent closed itemsets and the pseudo-closed itemsets which are defined in lattice theory, also define the correct association rules with a maximum confidence. In the basis, the rules are minimal antecedent and maximal consequent and non-redundant correct rules. As well as using the frequent closed itemsets, the proper basis and the structure bases for approximation association rules are defined by this work. A small set is the proper basis which containing the main informative and valuable approximation rules that are non-redundant informative rules.

All approximation rules are showed as summary of the structure bases that hold and useful when the proper bases is large. This research work proposes fast search algorithms proposed for providing these bases. Generating the inducing bases is performed without any access to the dataset, with the help of set of frequent closed itemsets.
This algorithm is identifying the closed and pseudo-closed frequent itemsets. Though this algorithm is allowing for the support of itemsets and, it works mainly in databases.

The novel method for generating set of frequent itemsets and association rules presented in this work prove that frequent closed itemsets, extends the Apriori algorithm and these it discovering the maximal frequent itemsets to generate frequent closed itemsets.

It adapts the results from the context of association rules for exact and partial implications. This edition is based on the generating set, gives new novel approach using frequent closed itemset, for generating bases for exact and approximation association rules. This method proposed is efficient for both improving the usefulness of extracted association rules and decreasing the execution time of the association rule extraction.

As shown by experiments, the proposed process for extracting bases for discovering association rules does not require any transparency compared with the customary approaches.

This research work addresses the notion of basis for commonly exact and approximation association rules. Various algorithms and the consequent part presents algorithms computing for discovering frequent and frequent closed itemsets are illustrated with the bases for association rules from the frequent closed itemsets in connection with proposed algorithm RSDCA.
3.2 FREQUENT ITEMSETS AND ASSOCIATION RULES CONSTRUCTION

This research work presents the association rule outline based on the closure operators and connection, mainly it was introduced. Context in data mining is usually defined as $d = o, i$ and $r$. Here finite sets of objects defined as $o$ and $i$ respectively. A binary relation between objects and items are defined as $r$ subset of $o \times i$.

The reality in each couple denoted by $(o, i) \in r$. The object $o \in o$ is linked to the item $i \in I$. Based on the object system, the result of an SQL query is a data mining framework and it can be a relation, or a class. $d$ is mixture of data mining objects which is identified by the oid.

A. Let itemset $I$ be a subset of item from $D$.

The support count of the itemset $i$ in $D$ is: $\text{supp}(\text{itemset } I) = \frac{||g(\text{itemset } I)||}{||O||}$. Itemset $I$ \cite{50} is said to be frequent, if the support of $I$ in $D$ is at least minimal support ($\text{minsupp}$). The set $L$ of frequent itemsets in $D$ is: $L = \{ \text{itemset } I \subseteq \text{itemset } I \} \cap \text{supp (itemset } I) \geq \text{minsupp}\}$\cite{50}.

An Association rule is an implication among two itemsets, with the form itemset $I_1 \rightarrow$ itemset $I_2$ where itemset $I_1$; itemset $I_1 \subseteq$ itemset $I$, itemset $I_1$; itemset $I_2 \neq 0$ and itemset $I_1 \cap$ itemset $I_2 = 0$. itemset $I_1$ and itemset $I_2$ \cite{50} are called equally the antecedent and the consequent of the rules.

An association rules support $\text{supp}(r)$ and confidence $\text{conf}(r)$ of $r$: itemset $I_1 \rightarrow$ itemset $I_2$ are defined as follows:
In association rules, the support and confidence is larger than or equal to the minsupp and minconf thresholds respectively.

The set association rules defined D with given minsupp and minimal confidence thresholds as follows:

**Association Rule** = \{r: \text{itemset } I_1 \rightarrow \text{itemset } I_2 \rightarrow \text{itemset } I_1 \mid \text{itemset } I_1 \subseteq \text{itemset } I_2 \subseteq \text{itemset } I \wedge \text{support (itemset } I_2 \geq \text{minimal support} \wedge \text{confidence (r) \geq \text{minimal confidence}}\}. Consider confidence (r)=1 then r is called an exact association rule or implication rule, or else r is called as an approximation association rule [50]. D for minmal support = 2/5 and minimal confidence = 1/2 are used to directly extract the exact and approximation association rules.

Frequent itemsets and association rules are created and formed by frequent closed itemsets. After that, it defined the association rules and the correct and structure bases for approximation association rules, and then extends the part of research towards situation of association rules.

A. The *closure operators* is \( h = f \circ g \) in \( 2^I \) and \( h^I = g \circ f \) in \( 2^O \), and the notation used are \( f \circ g (I) = f(g(I)) \) and \( gof(o) = g(f(o)) \) [50]. Given the

<table>
<thead>
<tr>
<th>Support ((r)) = ( |g(\text{itemset } I_1 \cup \text{itemset } I_2)| / |0|, )</th>
</tr>
</thead>
<tbody>
<tr>
<td>confidence (f(r) = \text{support (itemset } I_1 \cup \text{itemset } I_2) / \text{support (itemset } I_2) )</td>
</tr>
</tbody>
</table>
set \((f, g)\), the following properties hold for all itemset \(I\), itemset \(I_1\), itemset \(I_2 \subseteq I\) and \(O, O_1, O_2 \subseteq O\) [50].

i. itemset \(I \subseteq h(\text{itemset } I)\)  \(O \subseteq h^1(O)\)

ii. \(h(h(\text{itemset } I))=h(\text{itemset } I)\)  \(h^1(h^1(O)) = h^1(O)\)

iii temset \(I_1 \subseteq \text{itemset } I_2 \rightarrow h(\text{itemset } I_1) \subseteq h(\text{itemset } I_2)\)

iv \(O_1 \subseteq O_2 \rightarrow h^1(O_1) \subseteq h^1(O_2)\)

B. A closed itemset is a \textit{itemset} \(I \subseteq I\) in \(D\), if \(h(I) = I\). If the support of \(I\) in \(D\) is at least \textit{min-sup} then the \textit{closed itemset} \(I\) is said to be frequent. The smallest \textit{closed itemset} containing an itemset \(I\) is \(h(I)\), the closure of \(I\)[50]. The \textit{frequent closed itemsets} \(FC\) in \(D\) is defined as follows:

\[
\text{Frequent Closed} = \{ \ \text{itemset } I \subseteq I' \mid \text{itemset } I = h(\text{itemset } I) \wedge \text{support (itemset } I) \geq \text{minimal support} \} \]

A A Maximal set of items are a \textit{frequent closed itemset}, which support is at least \textit{minsupp} is common to a set of objects. The context \{e\} for \textit{minsupp}=2/5 frequent closed itemsets are presented. The \textit{frequent closed itemset} are the \textit{itemset} \{b,c,e\} while it is the maximal set of items frequent to the objects \{b,c,e\}.
The itemset \{b,c\} is not a frequent closed itemset because it is not a maximal set of items common to some objects: the items \{b\} and \{c\} (objects) are in all objects in relation and also in relation with the item \{e\}.

After this, one can exhibit that the set of frequent closed itemsets with their support is the least set from which frequent itemsets with their support and association rules can be generated is called generating set. The support of an itemset \(I\) is equal to the support of the smallest closed itemset containing \(I\):

\[ \text{supp}(I) = \text{supp}(h(I)) \]

\[ M = \{ \text{itemset } I \in \text{lattice } L \mid \text{there exists itemset } I' \in \text{lattice } L \text{ where itemset } I \subseteq \text{itemset } I' \} \]

where \(M\) is the set of maximal frequent itemsets is identical to the set of maximal frequent closed itemsets

\[ \text{MC} = \{ I \in \text{FC} \mid \exists I' \in \text{FC} \text{ where } I \subseteq I' \}. \]

<table>
<thead>
<tr>
<th>Frequent closed itemset</th>
<th>Support</th>
</tr>
</thead>
<tbody>
<tr>
<td>{}</td>
<td>5/5</td>
</tr>
<tr>
<td>{3}</td>
<td>4/5</td>
</tr>
<tr>
<td>{1,3}</td>
<td>3/5</td>
</tr>
<tr>
<td>{2,5}</td>
<td>4/5</td>
</tr>
<tr>
<td>{2,3,5}</td>
<td>3/5</td>
</tr>
<tr>
<td>{1,2,3,5}</td>
<td>2/5</td>
</tr>
</tbody>
</table>

Table 3.1: FC Itemsets Extract from D for minimal support
C. For all association rules holding in the dataset, the set FC of frequent closed itemsets with their support is a *generating set* for all frequent itemsets and their support. The maximal frequent closed itemsets are used to derive all frequent itemsets. The support of frequent closed itemsets is used to derive the support of each frequent itemset. After that, for both the set of frequent itemsets L and the set of association rules, FC (the set of frequent closed itemsets) is a generating set.

### 3.3 E(X)ACT ASSOCIATION RULES ALGORITHM (E(X)ARA)

The set of frequent - closed itemsets in P is FP. The basis for all correct association rule is set \( P = \{ r : \text{itemset } I_1 \rightarrow \text{itemset } (I_1) \cdot \text{itemset } I_1 \subset \text{itemset } I_1 \subset \text{frequent pattern} \uparrow \text{itemset } I_1 \neq \emptyset \} \) and also a binding in the dataset. While there no complete set with fewer rules than minimal with respect to the number of rules there are frequent - closed itemsets.

A frequent -closed *itemset I* is a frequent non-closed itemset is includes the closures of all frequent pseudo-closed itemsets[50]. The set of Frequent Pattern (FP) and frequently occurred pseudo closed itemsets and the related set of association rules are extracted from minimal support level of (minsupp is equal to 2/5) and minimum confidence level of (minconf equal to 1/2).

While the itemset \{a,c\} and the set \{b,e\} are not included in the itemset \{a,b\} and itemset \{a,b,c,e\}, itemset \{a,b\} is not consider as a frequent pseudo closed itemset closures of set \{a\}, set \{b\} respectively [50].
<table>
<thead>
<tr>
<th>Frequent pseudo-closed itemset</th>
<th>Support</th>
</tr>
</thead>
<tbody>
<tr>
<td>{a}</td>
<td>3/5</td>
</tr>
<tr>
<td>{b}</td>
<td>4/5</td>
</tr>
<tr>
<td>{c}</td>
<td>4/5</td>
</tr>
</tbody>
</table>

**Table 3.2:** Frequent Closed itemsets

<table>
<thead>
<tr>
<th>Exact rule</th>
<th>Support</th>
</tr>
</thead>
<tbody>
<tr>
<td>{a} → {c}</td>
<td>3/5</td>
</tr>
<tr>
<td>{b} → {e}</td>
<td>4/5</td>
</tr>
<tr>
<td>{e} → {b}</td>
<td>4/5</td>
</tr>
</tbody>
</table>

**Table 3.3:** Extract from minsupp

Also, Frequent Closed is the least developing a set for L and association rule (AR). Hence, steps over the dataset are greatly important to calculate the frequent itemset supports, the frequent itemsets could be derived from the most occurred frequent itemsets.

### 3.4 APPROXIMATION ASSOCIATION RULES

Let consider the frequent closed be the set of frequent closed itemsets [50]. The set prober basis = \{r: itemset I_1 → itemset I_2 → itemset I_1 \itemset I_1, itemset I_2 \itemset I_1, Frequent Closed \itemset I_1 \itemset 0 \itemset I_1 \itemset I_1 \itemset I_1 \itemset I_1 \itemset I_1 \itemset I_2 \itemset confidence (r) \itemset minimal confidence f\} [50] is a basis for all approximation association rules. Association rules in PB are proper approximation association rules. The proper basis for approximation association rules extracted from \{4\} for minimal support minsupp=2/5 and minimal confident minconf =1/2 are presented.
<table>
<thead>
<tr>
<th>Approximation rule</th>
<th>Support</th>
<th>Confidence</th>
</tr>
</thead>
<tbody>
<tr>
<td>{b,c,e} → {a}</td>
<td>2/5</td>
<td>2/5</td>
</tr>
<tr>
<td>{a,c} → {b,e}</td>
<td>2/5</td>
<td>2/3</td>
</tr>
<tr>
<td>{b,e} → {a,c}</td>
<td>2/5</td>
<td>2/4</td>
</tr>
<tr>
<td>{b,e} → {c}</td>
<td>3/5</td>
<td>3/4</td>
</tr>
<tr>
<td>{c} → {a,b,e}</td>
<td>2/5</td>
<td>2/4</td>
</tr>
<tr>
<td>{c} → {b,e}</td>
<td>3/5</td>
<td>3/4</td>
</tr>
<tr>
<td>{c} → {a}</td>
<td>3/5</td>
<td>3/4</td>
</tr>
</tbody>
</table>

Table 3.4: Proper Basis Extracted from \{d\} for \text{minsupp} = 2/5 and \text{minconf} = 1/2.

### 3.5 STRUCTURE BASIS FOR APPROXIMATION ASSOCIATION RULES

Let set of Frequent close itemset \text{FC} in D. And \( G_{\text{FC}} = (\text{vertices } V, \text{edges } E) \) is an undirected graph connected with \text{FC} where the set of vertices \( V \) and the set of edges \( E \). Consider \( F_{\text{FC}} = (\text{vertices } V, \text{edges } E) \) is the maximal confidence spanning forest connected with \text{FC}. The undirected graph is used to obtain the \( F_{\text{FC}} \) [50].

\( G_{\text{FC}} = (\text{vertices } V, \text{edges } E) \) is a suppressing transitive edges and cycles. While deleting some edges from the graph Cycles are removed and it comes in the last vertex I of the cycle. In this way all edges entering in itemset I.
After that, confidence which is smaller than the maximal confidence connected with an edge with the form \((\text{itemset } I_1, \text{itemset } I) \in E\) are finally deleted.

Let \(SB\) be the set of association rules represented by edges in FFC except rules from the vertex \(\{\emptyset\}\). The set \(SB = \{r : \text{itemset } I_1 \rightarrow \text{itemset } I_2 \text{ vertices } V \cup \text{itemset } I_1 \subseteq \text{itemset } I_2 \rightarrow \text{itemset } I_1 \neq \emptyset \cup (\text{itemset } I_1, \text{itemset } I_2) \in \text{edges } E'\}\) [50] is a basis for all approximation association rules holding in the dataset itemset \(I\) is the consequent of at most one approximation association rule in structure bases.[50].

The structure basis (SB) for approximation association rules extracted from \(\{d\}\) for minimal support (minsupp)=2/5 and minimal confidence (minconf) =1/2 is presented [50].

<table>
<thead>
<tr>
<th>Approximation rule</th>
<th>Support</th>
<th>Confidence</th>
</tr>
</thead>
<tbody>
<tr>
<td>{a,c} \rightarrow {b,e}</td>
<td>2/5</td>
<td>2/3</td>
</tr>
<tr>
<td>{b,e} \rightarrow {c}</td>
<td>3/5</td>
<td>3/4</td>
</tr>
<tr>
<td>{c} \rightarrow {a}</td>
<td>3/5</td>
<td>3/4</td>
</tr>
</tbody>
</table>

Table 3.5: Structure Basis Extract from \(\{d\}\) for minsupp = 2/5 and minconf =1/2.
3.6 GENERATING ASSOCIATION RULES

The code generating for exact association rules is re-presented in E(x)act Association Rule based Algorithm (E(X)ARA) figure.3.1 [50]. The $L_i, 1 \leq i \leq k$ are the input of the algorithm which containing the frequent itemsets and their support, and the sets $EFC_i, 0 \leq i \leq k$ is the frequent closed itemsets and their support. This algorithm first construct the frequent closed itemsets and then generate the association rules by the frequent closed itemset.

Initially, the set (EXA) exact association rule is set to the null set. If the null itemset is not a closed itemset, it is then certainly a closed itemset, and then it is added in $EFP_0$. If not $EFP_0$ is empty. Then, the algorithm reputedly recognized which i-itemsets in $L_i$ are closed from $L_1$ to $L_k$. in every iteration the set $EFP_i$ is set with the list of frequent i-itemsets that are not closed (step 5) and every frequent i-itemsets $l$ in $EFP_i$ is treated as follows.

The variable pseudo is initiallised to true. For each, if $p$ hold $l$ proves frequent closed itemset $p$ formerly discovered. Also, if the closure of $p$ is not holding $l$, then it is not closed and it is deleted from $FPI$. Or else, the closure minimum frequent closed itemset containing $l$ is determined. All the rules with the form $r : p$ gives $p$ are generated , when all frequent closed itemsets $p$ and their closure are fined once.. The algorithm in the set EXA contains all rules in the basis for exact association rules.
A. Correctness:

As the itemset $\phi$ has no subset and if it is not a closed itemset then it is by account a pseudo-closed itemset and the calculation of the set $FP_0$ is correct. The rightness of the calculation of frequent closed $i$-itemsets in $FP_i$ for $1 \leq i \leq k$.

If all frequent $i$-itemsets $l$ in $Li$ are not closed then $EFC_i$, are considered. That $l$ holds the closures of all frequent closed itemsets and that are subsets of $Li$, inserted in $EFPi$. If these $i$-itemsets are frequent closed $i$-itemsets and then the sets $EFPi$ are correct. All rules with a frequent closed itemset $I$ through association rules and generated in the final phase of the algorithm.

```
1) EXAR ← {}; // EXAR → Exact Association rules.
2) if (EXFC_0 = {}) then EXFP_0 ← {φ};
3) else EXFP_0 ← {};
4) for (i ← 1; i ≤ k; i++) do begin
5)   EFPi ← Li \ EFCi; // Li-Set of frequent $i$-itemsets and their support.
6)   for all itemsets $l$ ∈ EFPi do begin
7)     pseudo true;
8)   for all itemsets $p$ ∈ EFP_j with $j < i$ do begin
9)     if ($p \supseteq l$) and ($p$.closure $\subseteq$ 1)
10)    then do begin
11)   pseudo false;
12)   EFPi ← EFPi \ {l}; // EFPi- Set of frequent pseudo-closed $i$-itemsets, their closure and their support
13) endif
```
14) end
15) if (pseudo = true) then l.closure Min ⊆ (fc ∈ EFC_{j<i} | l ⊆ c));
16) end
17) end
18) forall sets EFPi where EFPi ≠ {} do begin
19) forall pseudo-closed itemsets p ∈ EFPi do begin
20) EXAR ← EXAR U {r : p ⇒ (p.closure − p), p.support};
21) end
22) end

**Figure 3.1:** E(X)ARA (Exact Association Rule bases Algorithm)

### 3.6.1 Generating Proper Basis for Approximation Association Rules

The code generating the proper basis for approximation association rules is well-known and Notations are given. The algorithm in figure 3.4 takes the sets $EFC_i, 1 ≤ i ≤ k$ as the input sets, hold the frequent closed non-empty itemsets and their support. The Proper Basis Approximation Association Rule (PBAA) is the output of the algorithm.

First (Proper basis for approximation association rules) PBAA is initialized to the null set. After that, the algorithm continuously considers all frequent closed itemsets $l ∈ EFC_i$ for $i ≤ k$. It determines frequent closed itemsets $l' ∈ EFC_{j<i}$ which are subsets of $l$ and calculates the association rules with the form $l' → l → l'$ that have sufficient confidence.

In, each itemset $l$ in $EFC_i$ is considered in the $i$th iteration. In $EFC_j$, $1 ≤ j < i$, a set $S_j$ it create all frequent closed $j$-itemsets that are subsets of $l$. 

72
Then, compute the confidence of the proper approximation association rule \( r : l' \rightarrow l - l' \) for each of these subsets \( l' \in S_j \).

In every iteration \( r \) is inserted in \( PBAA \), if the confidence of \( r \) is sufficient. The set \( PBAA \) contains all rules of the proper basis for approximation association rules, at the end of the algorithm.

Consider a subset function that takes a set \( X \) of itemsets and an itemset \( y \) as arguments and also it finds all itemsets \( x \in X \), that is subsets of \( y \).

In order to expand efficiency of the subset search, frequent and frequent closed itemsets are stored in a prefix-tree structure that is explained in the algorithm.

The algorithm scans all the correct approximation association rules, accordingly binding in the dataset for the exactness of the respective algorithm. The discussed algorithm computes, its subsets, amongst all other frequent closed item sets. Then, the brief confidence is ensured in the generation of all rules between two frequent closed itemsets.

The result of the PBAA is the proper basis for approximation association rules and these rules are all proper approximation association rules holding in the dataset.
3.6.2 Generating Structure Basis for Approximation Association Rules

The figure 3.5., shows the code generation of the structure basis for approximation association rules. The sets EFCi (Set of frequent closed i-itemsets and their support), 1 ≤ i ≤ k, of frequent closed non-empty itemsets and their support are taken as the input of the diagram. The basis for approximation association rules SBA are generated by this algorithm and are represented by the maximal confidence spanning forest F_{FC} associated with
\[ \text{EFC} = \bigcup_{i=1}^{j=k} \text{EFC}_i \text{ (without the empty itemset).} \]

The SBA(A)R (Structure basis for approximation association rules) is a set and set to the null set. Then, the algorithm continuously considers all frequent closed itemsets \( l \in \text{EFC}_i \) for \( i \leq k \). It finds the sets which are frequent closed itemsets \( l' \in \text{EFC}_{j<i} \) and covered by \( l \), i.e. they are straight predecessors of \( l \), and then calculates the maximal confidence association rules with the form \( l \rightarrow l' \rightarrow l \). In the \( i^{th} \) phases, every itemset \( l \) in \( \text{EFC}_i \) is treated as follows. The \( CR(\text{Set of candidate approximation association rules.}) \) is a set of candidate association rules with \( l \) in the consequent and is set to the null set. The set \( S_j \) holds all frequent closed \( j \)-itemsets for \( 1 \leq j < i \), in \( \text{EFC}_j \) that are subsets of \( l \) and are created.

According to the size, these subsets are considered in decreasing order. Then, the confidence of the proper approximation association rule \( r : l' \subseteq l - l' \) is computed, for each of the subsets. The \( r \) is inserted, if the confidence of \( r \) is sufficient and all subsets are removed from \( S_n < j \). Because of this rules with the form \( l'' \rightarrow l - l'' \) with \( l'' \in S_n < j \) they are transitive proper approximation rules. At last the candidate proper approximation rules with \( l \) in the consequent are calculated.

The maximum confidence value \( \text{maxconf} \) of rules are resolute the primary rule through a confidence level is inserted. At the last part of the algorithm, the structure basis for approximation association rules contains sets even including the all rules.
- **Correctness**: The discussed algorithm considers all the association rules lattice \( l' \rightarrow lattice \ l - lattice \ l' \) with confidence \( \geq \) minimal confidence fixed between two frequent closed itemsets \( l \) & \( l' \), i.e. \( l \) faces \( l' \). Association Rules (AR) are all proper basis approximation association non transitive rules. It can be represented by the graph \( G,F \) and \( C \) without transitive edges. Moreover, between all rules among the form \( X \rightarrow l - X \), at this point, maintain only the first one by confidence identical to the maximal value confidence of rules [50].

\[
X \rightarrow l - X.
\]

```
1) \( SBA \leftarrow \{\}; \)
2) \( for \ (i \leftarrow 2; \ i \leq k; \ i++ \) do begin \)
3) \( forall \ itemsets \ l \in \ EFC_i \ do \ begin \)
4) \( ECR \leftarrow \{\}; \)
5) \( for \ (j \leftarrow i-1; \ j > 0; \ j-\) do begin \)
6) \( S_j \ Subsets(\EFC_j , l); \)
7) \( end \)
8) \( for \ (j \leftarrow i-1; \ j > 0; \ j-\) do begin \)
9) \( forall \ itemsets \ l' \in \ S_j \ do \ begin \)
10) \( confidence \ f(r) \leftarrow \ l.\support / l'.\support; \)
11) \( if \ (confidence \ (r) \geq \ minimal \ confidence \ f) \)
12) \( then \ ECR \leftarrow \ ECR \cup \{r : l' \rightarrow 1 - l', \ l.\support, \ conf(r)\}; \)
13) \( for \ (n \leftarrow j-1; \ n > 0; \ n-\) do begin \)
14) \( S_n \leftarrow \ S_n - \ Subsets(S_n, l'); \)
15) \( end \)
16) \( endif \)
17) \( end \)
18) \( end \)
19) \( if \ (ECR \neq \{\}) \ then \)
20) \( \ maximum \ confidence \leftarrow Max_{ECR}(\text{confidence}(r)); \)
21) \( \text{find first} \ \{r \in \ ECR \ | \ \text{confidence}(r) = \text{maximal} \)
```
Figure 3.3 SBA (A)RA (Structure Bases for Approximate Rules Algorithm)

Sum amount of approximation conditioned association rules are in the countable number of proper basis in the structure basis for the number of non-transitive rules. For example the set D contain rules \( \{c\} \rightarrow \{a\} \) and \( \{a,c\} \rightarrow \{b,e\} \) are foend and the rule \( \{c\} \rightarrow \{a,b,e\} \) which is clearly transitive. While the confidence is multiplied by the two confidences is less than then this is less interesting.

By reducing the extraction from the itemset to applying the non-transitive rules in a proper basis for making the approximation association rules are to be made considerably interesting. Such association rules are generated with the pruning strategy and candidate rules are inserted in structure basis (SB). Datasets are usually relatively average size of bases compared with the sets of all rules obtained. In the case of inadequately correlated data, no exact rule is generated and the proper basis for approximation rules contains all approximation association rules that are in hold. Hence, all frequent itemsets are frequent closed itemsets.
On the other side of the correlated data, the numbers of extracted association rules in bases are much smaller than the total number of association rules that are in hold. In the Dataset the execution time of the computation of all the rules are applicable. Execution times of the derivation of the exact rules and the proper basis for non-transitive approximation rules are not presented, since they are considered in identical manner.

3.7 EXPERIMENTAL RESULTS

In this research, the re-presented algorithms [50] such as E(X)ARA, PBAA, SBARA are analyzed bases for association rules efficiently. A basis is a set of non-redundant rules from which all association rules captured the entire data it can be derived further from very large databases, accordingly its size also is extensively minimized and compared with the set of all possible rules for the purposes of reducing redundancy towards unwanted rules are eliminated.

This research approach has a two type of advantages: on first one is the user is provided with a less significant set of resulting rules. It is easier to handle, and information of enhanced feature. Further processing period is minimized and compared with the discovering of mining association rules in the very large databases. Such results are proved with the various datasets and the corresponding values are plotted in the graph.

Integrating reduction methods templates can directly be used for extracting from the bases all association rules matching some user specified
patterns. Information in taxonomies associated with the dataset can also be integrated in the process as proposed for extracting bases for generalized multi-level association rules.

Integrating item constraints and statistical measures in the generation of bases requires further work. Functional and approximation dependencies algorithms presented in this research can be adapted to generate bases for functional and approximation dependencies.

The functional dependencies constituted of minimal non-trivial functional dependencies. Hence, the number of rules is minimal; moreover these rules have minimal antecedent and maximal consequent.

Furthermore, the proper and structure bases for approximation rules are also smaller than the basis for approximation dependencies defined. Adapting these algorithms is used to find the appropriate functional dependencies in voluminous databases.
Figure 3.4 Comparison Execution Time between FP-Growth, H-mine, RFP in various support level using Connect Dataset
Figure 3.5 Comparison Execution Time between FP-Growth, H-mine, RFP in various support level using Mushroom Dataset
Figure 3.6 Comparison Execution Time between FP-Growth, H-mine, RFP in various support level using pumsb Dataset
The following figures 3.7, 3.8 and 3.9 illustrated the comparison of exact time obtained by implement the proposed algorithms namely $E(X)ARA$, $PBAA-RA$, and $SBA(A)RA$ with support level used to calculate the execution time testified through Connect, Chess and mushroom datasets.

**Figure 3.7** Comparison Execution Time between $E(X)ARA$, $PBAA-RA$ and $SBA(A)RA$ in various support levels using Connect Dataset
Figure 3.8: Comparison Execution Time between E(X)ARA, PBAA-RA and SBA(A)RA in various support levels using Chess Dataset
Figure 3.9  Comparison Execution Time between E(X)ARA, PBAAR-RA, and SBA(A)RA in various support levels using mushroom Dataset
3.9 SUMMARY

In this research, the re-presented algorithms such as $E(X)ARA$, $PBAA-R\!A$, and $SBA(A)RA$ are compared bases for association rules efficiently with respect to presented algorithm RSDCA for the purpose of finding optimum discussion on the area of pattern growth methods that is key functional aspects in large databases research methods, later involved in this research and developed RSDCA algorithm. A basis is a set of non redundant association rules that are significantly derived and captured related information from datasets. Additionally, its size also considerably reduced and compared with the set of all possible rules and redundant, useless rules are accordingly discarded. This research approach has dual advantages: first advantage is, the user is given with a minimal set of resultant association rules, it is easier to handle, and further information is of higher standard. the second advantage is, the processing times are decreased compared with the finding of all mining rules. Such results are proved with the various datasets and the corresponding values are plotted in the graph. Integrating reduction methods Templates can directly be used for extracting from the bases all association rules matching some user specified patterns. Information in taxonomies associated with the dataset can also be integrated in the process as proposed for extracting bases for generalized multi-level association rules. Integrating item constraints and statistical measures in the generation of bases requires further work. Functional and approximation dependencies algorithms presented in this research can be adapted to generate bases for functional and approximation
dependencies. The functional dependencies constituted of minimal non-trivial functional dependencies. Hence, the number of rules is minimal; moreover these rules have minimal antecedent and maximal consequent. Furthermore, the proper and structure bases for approximation rules are also smaller than the basis for approximation dependencies defined. Adapting these algorithms is used to find the appropriate functional dependencies in voluminous databases.