ISSN (Print) : 0974-6846
Indian Journal of Science and Technology, *Vol 8(31), DOI: 10.17485/ijst/2015/v8i31/76281, November 2015* ISSN (Online) : 0974-5645

Exceptional Patterns with Clustering Items in Multiple Databases

R. Suganthi[1*] and P. Kamalakannan[2]

[1]Department of Computer Applications, Valluvar College of Science and Management, Karur - 639003, Tamil Nadu, India; mcasuganthi2011@gmail.com

[2]Department of Computer Science, Arignar Anna Government Arts College, Namakkal - 637 001, Tamil Nadu, India; kamal_karthi96@yahoo.co.in

Abstract

The information from multiple local databases can be mined together to make global patterns. More global decisions will be based on synthesized patterns and exceptional patterns using clustering technique. To take better decision in organization head quarter level, exceptional patterns also need to be analyzed for non profitable things which help for continuous company growth. Our new strategy is developed considering both clustering the frequent items also exceptional patterns. Various experiments are conducted with 10 sample datasets and the results are recorded in experimental section, with this the significance and limitations behind those approaches can be made clear.

Keywords: Exceptional Patterns, Frequency Item Sets, Global Patterns, Multi Database Mining

1. Introduction

The proposed strategy supersedes the existing traditional data mining techniques based on data warehousing where the huge investment has to be made on software and hardware. Also, it is very complicated to fetch the required data. In order to get the exact frequent items sets and the exceptional item sets from the multiple databases. The following steps are used to finding the different types of patterns which is major role in decision making. Traditional data mining techniques, measures of association[1], frequent items in association mining, synthesizing support of an item set, clustering local frequency items[2], and identifying exceptional patterns. With the help of the all above considerations, various patterns are discovered through the designed algorithm. Clustering the frequent items and finding exceptional patterns are the vital role in head office to take decision in proper manner which leads the company for the rapid progress of the organization. Knowledge discovery from multiple databases known as the concept of Multi Database Mining (MDM). It can be defined as the process of mining data from multiple databases, which may be heterogeneous and finding novel and useful patterns of significance[3]. Without moving the data to the central repository, the beginning step of the local pattern analysis in multi data base accomplishes the different data mining operations which is depends on the distributed resources and data types.

A local pattern would be a frequent item set which pertaining to show the individual local database. This local pattern analysis is suggested when the application involves a huge number of data sources. But Adhikari et al.[4] denounce the issue, frequency of data mining is a weak spot of local pattern analysis approach for MDM problem. The group of frequent item sets wraps up major uniqueness of a database. Many appealing algorithms[5–7] have been projected for mining frequent item sets in a database. Consequently there are many implementations[8] for extort frequent item sets. The first step is traditional

** Author for correspondence*

data mining techniques are used to do the preprocessing work in all databases. Item set are highly associated, if one of the item of 'P' is purchased than the remaining items of 'Q' are also possible to be purchased in the similar way of the transactions. This needs to enter into frequent item in association mining. The finding of frequent items in given database is elaborately given in the Tables 1, 2, 3, 4 and 5. Rules are framed according to those items. After synthesizing, the important process is filtering the irrelevant items from all the branch local databases. With the synthesized support of all item sets in multiple databases, the major data mining techniques, clustering frequency items and exceptional patterns are to be implemented through the designed algorithm. During the decision making process in head quarters, they have to analyze the different types of patterns which would be given profit in all ways or not. The producing of frequent items using clustering techniques will provide the relationship among the items which symbolically indicates that what are all the items are mostly purchased together in day to day life. And also provides the details about what kind of products are moving fast and also it explore the role of sales manager decision in all aspects. When working with the designed algorithm to produce the frequent items and exceptional patterns, it steer the clear patterns usage and through the algorithm. This would be considered to avoiding the more effort on products like transportation cost, human effort and expiry of the products using this kind of patterns. Based on these two types company can come into clear strategy about the products in all ways.

2. Associations

Basically extracts the patterns from the database, based on the 2 measures such as minimum support and minimum confidence. These 2 types of measures are important for mining frequent item set mining and association rule generation. Frequent patterns are patterns that appear frequently in a dataset. An emblematic case of frequent item set mining is Market Basket analysis[9]. This process analysis the habits of customer buying throw the finding of association between the different items. For example *jeans* and *white shirt* may have often been purchased at one time from a department store and black trousers and blue T-shirt often purchased at another time. This worthy information can direct to increase the sales by selecting customer information and helping retailers do selective marketing and plan their shelf space. The association rule represents,

Computer--> antivirus software (support = 2%, confidence = 60%).

Here the support measure 2% means that computer and antivirus software are purchased together. A confidence of 60% means that 60% of the customers who purchased a computer also bought the software by parallel.

2 important steps in association rule mining:

- Making all items sets having support factor greater than (or) equal to the user particular min support.
- Making all rules having the confidence factor greater than (or) equal to the user defined min confidence.

To demonstrate the use of the support-confidence frame work, we detail the process of mining association rules by an example as follows.

Let the item universe be I = {D1, D2, D3, D4, D5}; and a transaction database be TID = {100, 200, 300, 400, 500}. The data in the transactions are listed in Table 1.

Table 1. Model transaction database

Transaction ID (TID)	Data base - items					
500	P		r	s		
600		q	r		t	
700	P	q	r		t	
800		q			t	u

In Table 1: 500, 600, 700, 800, 900 are the distinctive identifiers of the 4 transactions and

p = Bathing Bar
q = Shampoo
r = Washing Bar
s = Detergent powder
t = Hand wash
u = Floor cleaner

Here 4 rows contain 6 data items. We can discover association rules from these transactions using the support and confidence framework.

Using this example, 2 step association rule mining[10] as follows:

- **The first step is to count the frequencies of 'k' item sets.**

In Table 1, item[p] take place in 2 transactions.

TID = 500 and TID = 700 its frequency is 2 and its support (supp(p)) is 50% which is equal to min supp = 50%, item{Q} occurs in 3 transactions.

TID = 600,700 and 800 its frequency is 3 and its support supp (q) is 75% which is larger than min supp; item{r} take place in 3 transactions TID = 500, 600, 700, its frequency is 3 and its support supp (q) is 75% which is greater than min supp; item{s} take place in 1 transactions TID = 500, its frequency is 1 and its support supp{s} is 25% which is less than min supp; item{t} occurs in 3 transactions TID = 600,700,800; its frequency is 3 and its support supp(t) is 75% which s greater than min supp. Item{u} take place in 1 transactions TID = 900, its frequency is 3, its support supp{u} is 25% which is fewer than min supp. They are summarized in Table 2.

Table 2. Single item sets in the database

Item sets	Frequency	>Min Support.
{p}	2	yes
{q}	3	yes
{r}	3	yes
{s}	1	no
{t}	3	yes
{u}	1	no

Table 3. Two item sets in the database

Item sets	Frequency	>Min Support
{p,q}	1	no
{p,r}	2	yes
{p,s}	1	no
{p,t}	1	no
{q,r}	2	yes
{q,t}	3	yes
{q,u}	1	no
{r,s}	1	no
{r,t}	2	yes
{t,u}	1	no

We now consider 2 item sets for Table 2 Item set {p, q} occurs in 1 transaction TID = 700, its frequency is 1 and its support supp(p ∪ q) is 25% which is fewer than min supp = 50%. Item set {p, r} occurs in 2 transactions TID = 500 and TID = 700 its frequency is 2 and its support supp(p ∪ s) is 50%, which is equal to min supp = 50%, item set {p, t} take place in one transaction TID = 700, its frequency is 1 and support supp(p ∪ t) is 25%, which is fewer than min supp = 50%; item set {q, r} take place in 2 transactions TID = 600 and TID = 700 its frequency is 2 and its support supp(q ∪ r) is 50%, which is equivalent to min supp of 50%. Item set {q, t} occurs in 3 transactions TID = 600, 700, 800, its frequency is 3 and its support supp(q ∪ t) is 75%, which is greater than min supp. This is reviewed in Table 2 item sets in the database.

Also, we can obtain 3-item sets and 4-item sets as listed in Table 4 and 5.

Table 4. Three item sets in the database

Item sets	Frequency	>Min Support
{p,q,r}	1	no
{p,q,e}	1	no
{p,r,s}	1	no
{q,r,t}	2	yes
{q,t,u}	1	no

Table 5. Four item sets in the database

Item sets	Frequency	>Min Support
{p,q,r,t}	1	-

- **The second step is to produce all association rules from the frequent item sets.**

Because only one frequent item set is in Table 5.

4 item sets do not contribute any valid association rule[11]. In Table 1.4 there is one frequent item set {q,r,t} with supp {q ∪ r ∪ t} is 50% of min supp. For frequent item set {q, r, t} because supp (q ∪ r ∪ t)/supp(q ∪ r) = 2/2 = 100% is greater than min conf = 60%, q ∪ r-> t can be extracted as a valid rule; because supp (p ∪ q ∪ t)/supp(q ∪ t) = 2/3 = 66.7% is greater than minimum confidence, q ∪ t--> r can be extracted as a valid rule because supp(q ∪ r ∪ t)/supp (r ∪ t) = 2/2 = 100% is greater than minimum confidence, r ∪ t--> p can be extracted as a valid rule; and because supp (q ∪ r ∪ t)/supp (P) = 2/3 = 66.7 is greater than min conf, q-->q ∪ t can be extracted as a valid rule and so on. The generated association rules from {q,r,t} are in Table 6 and 7.

Example database

D1 -->{a, b} #{a, b, c} # {a, b, c, d} #{c, d, e} #{ c, d, f} #{c, d, i}

D2 -->{a, b} # {a, b, g} # {g}

D3--> {a, c, d} # {a, b, d} # {c, d}

D4 -->{a} # {a, b, c} # {c, d} # {c, d, i}...D(10) {a, c, d} # {a, b, d} # {c, d}

Assume that α = 0.5 and β = 0.7. Let X (n) represent the fact that the item set X has support (n) in the db, in the beginning the databases are sorted in non-increasing order on database sizes. The sorted database is given as follows:

[D1 = 20, D2 = 7, D3 = 10, D4 = 9, D = 11

D6 = 10, D7 = 9, D8 = 8, D9 = 8 D10 = 8]

Sorted databases (D1, D5, D6, D3, D4, D7, D9, D8, D10, D2)

By applying pipe lined feedback technique (PFT), we attain the following item set in diverse local database.

LPB (D1, α) = {{1}(1),{2}(0.999), {4}(0.657), {5}(0.83),{6}(0.978),{7}(1),{8}(0.998),{10}(0.561), {12}(0.998), b {14}(0.626),{15}(0.784),{16}(0.977),{17}(1),{18}(0.996), {1,2}(0.999),{1,4}(0.657),{1,5}(0.83),{1,6}(0.978),{1,7}(1), {1,8}(0.98),{1,1}0(0.561),{1,2,10}(0.56),{1,2,12}(0.998), {1,2,14}(0.626)

Let D be the union of D1, D2…….., D10

Synthesized HFIS in "D" are specified as follows: SHFIS (D,0.5,0.7) = {b, d} = 0.63,{b, f}(0.51) {b, h}(0.65), {c}(0.5), {d}(0.83),{d, h}(0.57) {4}(0.51)

The SHF1S>1, one forms the basis of the proposed clustering techniques.

Table 6. For frequent-3 item sets; start with 1 - item consequences

Rule No	Rule	Confidence	Support	>min conf
Rule 1	q-->R ∪ T	100%	50%	YES
Rule 2	r-->Q ∪ T	66.70%	50%	YES
Rule 3	q-->P ∪ R	100%	50%	YES

Table 7. Form all 2 item consequences from high conf 1 - item consequences

Rule No	Rule	Confidence	Support	>min conf
Rule 4	q-->R ∪ T	66.70%	50%	YES
Rule 5	r-->Q ∪ T	66.70%	50%	YES
Rule 6	t-->Q ∪ R	66.70%	50%	YES

Also, we can produce all association rules from frequent 2 item sets as shown in Table 3 they are illustrated in Tables 6-11 through the Tables 1-5.

Table 8. For 2 item sets; start with 1 item consequences for {P, R}

Rule No	Rule	Confidence	Support	>min conf
Rule 7	P-->R	100%	50%	YES
Rule 8	R-->P	66.70%	50%	YES

Table 9. For 2 item sets; start with 1-item consequences for {Q, R}

Rule No	Rule	Confidence	Support	>min conf
Rule 9	Q-->R	66.70%	50%	YES
Rule 10	R-->Q	66.70%	50%	YES

Table 10. For 2 item sets; start with 1 item consequences for {Q, T}

Rule No	Rule	Confidence	Support	>min conf
Rule 11	Q-->T	100%	75%	YES
Rule 12	T-->Q	100%	75%	YES

Table 11. For 2 item sets; start with 1 item consequences for {R, T}

Rule No	Rule	Confidence	Support	>min conf
Rule 13	R-->T	66.70%	50%	YES
Rule 14	T-->R	66.70%	50%	YES

The 14 association rules listed in the above data set can be extracted as valid rules according to the definitions.

The difficulty of mining association rules is to generate all rules p-->q that have both user specified thresholds greater than(or) equal to support and Confidence, called minimum support (min supp) and minimum confidence (min conf) respectively. For standard associations,

Supp (p ∪ q)> = min supp;

Conf (p--> q) = supp(p ∪ q) / supp(p)> = min conf.

These details explained more with the above descriptions and examples of frequent item sets.

Association analysis is discovered to cause rules and algorithms from large databases. For computation methods of association analysis can be found in association mining related journals.

In order to apply interesting association analysis, a wide range of problem has been examined over such diverse topics as models for discovering generalized association rules. In recent times, an FP based frequent patterns mining method was developed by Zhang et al[12]. This procedure direct to 3 benefits. It can condense a large database into a highly condensed, much smaller data structure, so as to avoid the expenditures of the company, scanning of repetitive databases.

- FP-tree based mining accepts a pattern fragment growth method to evade the costly generation of a huge number of candidate sets.

- The separation is based divide and conquer method can decrease the dimension of the subsequent conditional pattern bases and conditional FP-trees. This experiments shows that this technique is more efficient than the apriori algorithm[3]
- For finding the frequency item sets in multiple databases the first step is to resolve the similarity between each pair of database using the proposed measure of similarity algorithm based on the sets of frequent item sets in a pair of databases. One could define many measures of similarity between them. We propose 2 measure of similarity between a pair of databases. The first measure similr_1 is defined as follows:

2.1 Definition 1

The measure of similarity similr_1 between databases D1 and D2 is defined as follows.

$$similr_1(D1, D2, \alpha) = (|\ FIS(D1,\alpha) \cap FIS(D2,\alpha)\ |) / (|\ FIS(D1,\alpha) \cup FIS(D2,\alpha)\ |)$$

Here the symbols ∪ & ∪ stands for the intersection and union operations of set theory respectively. The similarity measure similr_1 is the ratio of the number of frequent item sets common to D1 and D2 and the total number of distinct frequent item sets in D1 and D2. Frequent item sets are the dominant patterns that establish the major characteristics of a database. Accordingly, a good measure of similarity among two databases is a function of the support of frequent item sets in the databases. Our second measure of similr_2 defined as follows:

2.2 Definition 2

The measure of similarity similr_2 among databases D1 and D2 is defined as follows:

$$Similr_2(\text{D1},\text{D2},\alpha) = \frac{\sum \min\{\text{supp}(\text{X},\text{D1}),\text{supp}(\text{X},\text{D2})\}\ \text{X} \in \{\text{FIS}(\text{D1},\alpha) \cap \text{FIS}(\text{D2},\alpha)\}}{\sum \max\{\text{supp}(\text{X},\text{D1}),\text{supp}(\text{X},\text{D2})\}\ \text{X} \in \{\text{FIS}(\text{D1},\alpha) \cup \text{FIS}(\text{D2},\alpha)\}}$$

Thus the similarity measures[13] among the items present a numerical estimate of statistical dependence among a set of items that is, items of interest ‘Y’ are highly associated, if one of the items of ‘Y’ is purchased then the remaining items of ‘X’ are also possible to be purchased in the same transaction. Agarwal et al.[14] have projected support measure in the context of finding association rules in a database to find support of an item set it requires frequency of the item set in a given database. An item set in a transaction could be a source of association among items in the item set. But support of an item set does not consider frequencies of its subsets. As a result, the support of an item set might not be a fine measure of association among items in an item set.

Hershberger and Fisher et al.[15] discussed about some measures of association proposed in statistics. Measures of association could be classified into two groups. Some measures compact with a set of objects, or could be generalized to deal with a set of objects. Cosine and correlation are used to measure the association between two objects. These measures might not a suitable as a measure of association among items of an item set. Confidence and conviction are used to measure strength of association between item sets in some sense. These measures might not be useful in the current context, since we are interested in capturing association among items of an item set.

3. Synthesizing Association among Items

The important issue in MDM is synthesizing support of an item set. In this paper, we present an algorithm for synthesize association among items in ‘D’ we discuss here various data structures required to implement the proposed algorithm. Let ’N’ be the number of frequent item sets in D1, D2…….Dn.

The array of item set extracted during mining multiple databases by using AIS (Array of Item Set) and Item Sets (IS) which is stores into the array of IS. It will be sorted based on the item set attribute the variable “total Trans” add sizes of all branch databases. For mining multiple databases algorithm pipelined feedback technique has been used. In order to improve the quality of synthesized global pattern[16], this technique has given better performance as compared to existing techniques. Here High Frequency Item Sets (HFIS) are important to take decision in management level during the synthesizing high frequency item set process in MDM; also we considered the exceptional patterns in the process. Because of these kind of patterns also provide valuable information to the headquarters in which the low sales of the non profitable products in the company which provides the easy solutions to plan to manufacture such items. It may be reduce the transportation cost and efforts of company. So we had taken 2 major roles in this. The first role is to clustering the local frequency items and finding the exceptional patterns which is described in following section.

4. Multi Database Mining Methods

There are two types in MDM.

- Generalized MDM method.
- Specialized MDM method.

These two methods could be used in variety of multi data base mining applications towards fulfilling their operational needs. Local pattern analysis, partition algorithm, identify exception pattern algorithm. Rule synthesizing algorithm, are comes under generalized MDM techniques. Each technique has some drawbacks to identify the patterns. So we are in need of specialized MDM techniques. For mining multiple real databases, Adhikari and Rao have projected association Rule Synthesis Algorithm (ARS) for synthesizing association rules in multiple real databases. The estimation procedure captures such trend and estimates the support of a missing association rule. Though, association rule synthesis algorithm might revisit the approximate global patterns. Due to that revisit of approximate global patterns, Adhikari and Rao overcome the problem with the help of most of mining global patterns of select items which is enlightened below:

- Each branch office constructs the forwarded databases sends it to the central office.
- Also, each branch extracts patterns from its local database.
- The central office clubs this forwarded database (patterns) into a single database.
- A traditional DM technique could be applied to extract patterns from forward database.
- The global patterns of select items could be extracted effectively from local patterns and the patterns extracted from forward and one more important MDM specialized technique is PFT, which progress the quality of synthesized patterns as well as an inspecting the local patterns significantly. Before applying PFT, need to prepare data ware houses for different branches of a multi branch organization. In PFT (Pipelined Feedback Technique), W, is mined using a SDMT (single data mining technique) and local pattern base LPB, is extracted.

Thus, |LPB i-1|<=|LPB i-1| <=|LPB i|, for i = 2, 3, .., n.

These are n! arrangements of pipelining for n databases. For the purpose of growing number of local patterns, Wi precedes Wi-1 in the pipelined arrangement of mining data ware houses if size (Wi) > = size (Wi-1). For i = 2, 3,….. ,n. Finally we study the patterns in LPB1, LPB2, LPB3, and LPB n for synthesizing global patterns (or) analyzing local patterns. Again Adhikari[17] proposed technique for significant exceptional patterns in MDM which represents the global exceptional frequent item sets. The average of support can be obtained by the following formula:

$$\text{avg}\left(\text{supp}(\text{x}), \text{D1}, \text{D2...D}_\text{k}\right) = \sum_{i=1}^{k} \text{supp}\ \alpha(\text{X}, \text{Di})(\text{X}, \text{D2}) / \text{k}$$

Here, Di-exceptional source with respect to the global exceptional frequent item set X, if supp (X,Di) > = average (supp(X),D1, D2,…., D_n) i = 1, 2, …,k.

After sorting the local databases, we could obtained the different local database item sets which are useful to synthesizing the high frequency item set in multiple data bases. During synthesizing process two types of patterns are considered. These two patterns (synthesized patterns and exceptional patterns) are important to take decision in head quarters level.

Based on support and thres hold values, it provides the frequency of items in all local databases. Here all items are not comes under the category of synthesized patterns, some patterns would not be matched with criteria, that type of patterns are called exceptional patterns and a pattern that has been mine from most of the databases but has a small support or high support in databases 'D'. Both the types of exceptional patterns are global in nature, since all the branch databases are well thought-out in this paper; we are concerned in mining synthesized high frequency patterns and exceptional patterns.

Algorithm: Rule of Synthesizing and Exceptional Patterns

Input: DB-Database

n – Number of Database
min sup - Minimum Support
thre – Threshold Value
μ **-** Exceptional Threshold Value

Output:

Exceptional Pattern
SHFIS
Clustering Local Frequency Items

Steps:

- Store every one of the local item sets into array IS;
- Arrange item sets of IS
- Find GP (Global Pattern) from IS
- Find avg = IS size/GP Size
- For each Pattern P in GP do
- Count Num(P) in IS
- if ((Num(P) < avg) then
- Add P to CEP (Candidate Exceptional Pattern)
- End if

- End for
- For each candidate exceptional pattern P in CEP do
- Find

$$Supp_G(P) = \frac{\sum_{1}^{Num(P)} \frac{Supp_i(P) - \min \sup}{1 - \min \sup}}{Num(P)}$$

- If $Supp_G(P) > \mu$ then
- Add P to EP (Exceptional Pattern)
- End if
- End for
- Let nSynItemSets = 0; let i = 0;
- While (i < |IS| -1) do
- let j = i; let count = 0;
- While (j ≤ i+n and j < |IS|-1) do
- if (IS(j).itemset = IS(i).itemset) then
- count++
- j ++;
- End if
- End While
- If (count/n > thres) then
- Add Item Set to SIS (Synthesized Item Set)
- End if
- i = j
- nSynItemSets++;
- End While
- Sort SIS with non-increasing order
- If (m==1) then (m : number of SHFISs of size larger than one)
- Form a single class
- End if
- mutualExcl=false; temp=null
- for i =1 to m do
- if temp! = null
- mutualExcl = true;
- End if
- if ((mutualExcl = false) && (SA(S(i), D) != SA(S(i + 1), D))) then
- for j = 1to i do
- create a class using items in S(j);
- temp = temp ∪ S(j);
- End for
- End if
- End for

5. Experiments

In order to verify the efficacy of our proposed method we carry out some experiments we chose a databases 'D1……..D10' that is available in http//fimi.ua.ac.be/data/ and this analysis the synthesis process among all the databases. The dense data set contains distinct items for MDM, the database is divided into 10 data sets, namely D1,……D10 with a transaction 10,000 respectively.

We have taken several experiments to examine the effectiveness of our approach. All the experiments have been implemented on a 2.8 GHz Pentium D dual processor with 512 MB of memory using Net Beans IDE (version 8.0) software. We present the experimental results using 10 databases via retail and mushroom. Each database contains 10,000 records respectively. We have estimate the success of our synthesizing approach by conducting various experiments. The characteristics of these databases are discussed further.

Study 1

Table 12. Database characteristics

Data base (DB)	Number of Transaction (NT)	Average Length of Transactions (ALT)	Average Frequency Items (AFI)	Number of Items (NI)
D1	10,000	11.05	127.655	820
D2	10,000	11.11	128.031	862
D3	10,000	11.03	127.044	865
D4	10,000	11.12	127.043	866
D5	10,000	11.12	128.064	864
D6	10,000	11.13	128.034	862
D7	10,000	11.09	128.026	860
D8	10,000	11.09	128.051	861
D9	10,000	11.08	128.050	864
D10	10,000	11.07	128.000	865

We have synthesized database of 10,000 records with 20 items and the entire database divided into 10 databases. Transactions and items of the 10 databases are varied. For D1 and D2 transactions, Ex, D1 = {a, b},{a, b, g},{c, d },{a}. In the initial stage, 10 databases are sorted according to their size. Then the local patters are obtained. If α = 0.5 and β = 0.7, then the set of item sets are produced by synthesized algorithm. Comparison of synthesized global support and confidence values are given in Table 14 and Table 15 presents a different error measures.

Study 2

For the second study, a database having 10,000 records with different number of items. When producing the synthesized high frequency items, some item sets are not comes under the category of high vote patterns, such patterns are considered that the comparison of synthesized high frequency item sets support and confidence values and error measures are displayed in Table 13.

Table 13. Synthesizing time with all database Average Errors

S. No	Alpha	Gamma	Average Error	Synthesizing Time
1	0.1	0.7	0.0400320	2 minutes
2	0.2	0.7	0.0728418	29 Seconds
3	0.3	0.7	0.0735452	21 Seconds
4	0.4	0.7	0.7528169	15 Seconds
5	0.5	0.7	0.0661811	12 Seconds
6	0.6	0.7	0.0696122	9 Seconds
7	0.7	0.7	0.0849281	2 Seconds
8	0.8	0.7	0.0849281	2 Seconds
9	0.9	0.7	0.0870122	1 Second
10	0.1	0.7	0	0 Seconds

Study 3

For the third study, the first databases are the same as for study 1 with 10,000 records. Comparison of synthesized support and confidence values are specified and error measures are considered in Table 13. Exceptional patterns and synthesized high vote pattern given in Table 14.

6. Analysis of Results

The proposed method always furnishes the way of correct results of synthesized patterns and exceptional patterns with the different support and confidence, values[18].

We have experiment the number of synthesized and exceptional patterns in multiple databases at different α and β values[19]. We present experimental results in Figures 1, 2, 3 and 4.

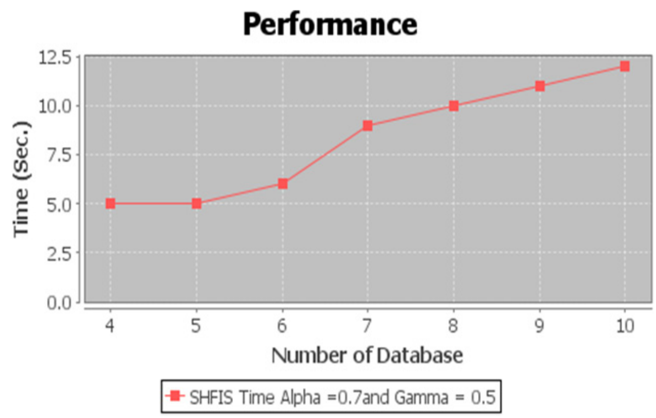

Figure 1. Number of patterns performance and its time.

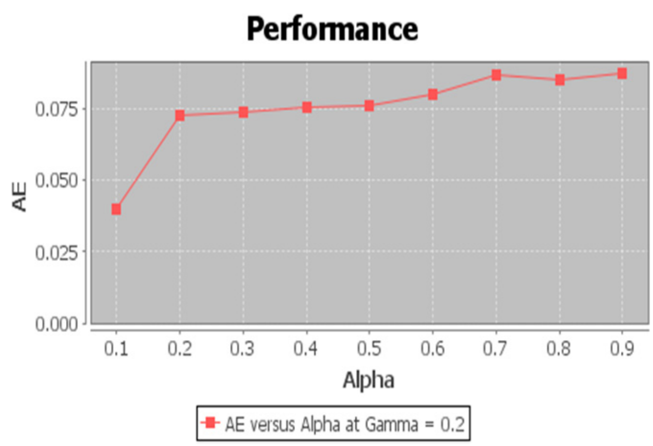

Figure 2. Number of patterns performance and its time.

Table 14. Synthesized and exceptional patterns

S. No	Alpha	Gamma	Exceptional Patterns	Synthesized Patterns	Local frequency items
1	0.5	0.7	NIL	a:0.5835, c:0.5625, c,d:0.4585, d:0.54175	a,b,g,{c,d}
2	0.2	0.7	{a,d}:58375,g:58375	a:0.5875,a,b:0.4375,a,c:0.2289,b:0.4375,c:0.5625,c,d:0.4585,d:0.5417	{a,b,g,i,{c,d}}, {{g},{i},{c,d},{a,d}}, {{g},{i},{{c,d},{a,b},{a,c}}
3	0.3	0.7	{a,d}:0.5242,{g}:0.5242	a:0.5835,a,b:0.375,b:0.375,c:0.5625,c,d:0.4585,d:0.54175	{{a},{b},{g},{c,d}},{g},{c,d},{a,b}
4	0.4	0.7	{a,d}:0.4450,{g}:0.4450	a:0.5835, c:0.5625,c,d:0.4585,d:0.54175	a,b,g,{c,d}
5	0.1	0.7	{g}:0.63	a:0,5835,a,b:0.4375,a,c:0.22899,b:0.4375,c:0.5625,c,d:0.4585,d:0.54175	{{a},{b},{e},{f},{f},{g},{i},{c,d},{{e},{f},{g},{i},{c,d},{a,b}},{{e},{f},{g},{i},{c,d},{c,d},{a,c}}

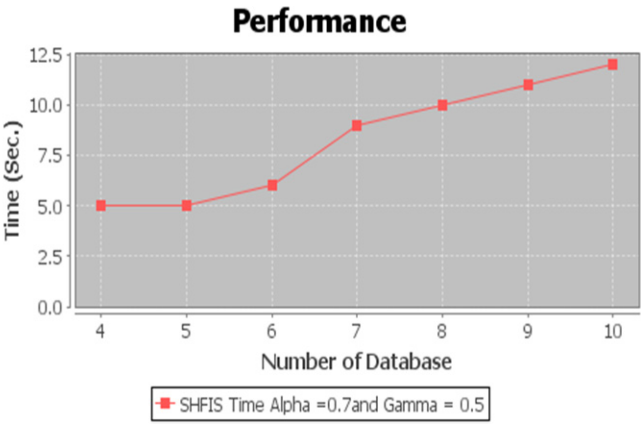

Figure 3. Number of patterns performance and its time.

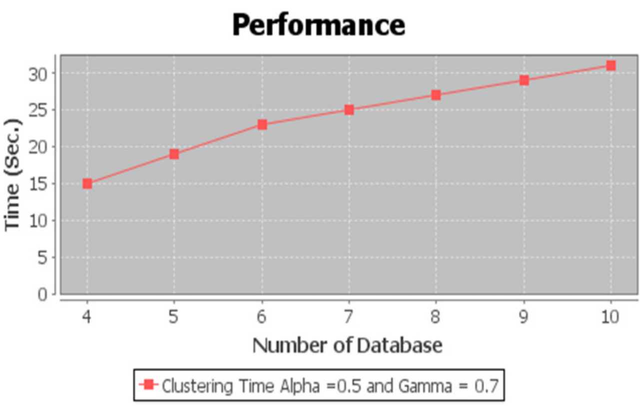

Figure 4. Number of patterns performance and its time.

6.1 Error of an Experiment

To assess the proposed technique of synthesizing clustering frequent items with exceptional patterns, we have considered the quantity of in accuracy occurred in an experiment. There are many algorithms[20,21] for mining exceptional and synthesized patterns separately. Error of the experiment is comparative to the number of transactions (i.e., the size of the database, number of items, and length of the transaction in a database. Thus the error of the experiment needs to be expressed with the ANT, ALT and ANI in a database where ANT, ALT and ANI stands for the average number of transactions, average length of transaction, and the average number of items in a database respectively. The error of the experiment is based on the clustering frequent items and exceptional patterns in database D. We have calculated the error of the experiments in Table 13 and it mention the AE_S (Average Errors) at different α β, values which is used to study the relationship between them. Experimental results are given in the format of graphs in 1, 2, 3 and 5.

If alpha (α), beta values (β), at support level 0.5 and 0.7 ,all item set become frequent. Thus the average error of an experiment is 0 when the alpha and beta values 0.5 and 0.1 respectively. The above figures show small differences between the AEs. During the decision making time in the head quarter level, various patterns will be considered based on the criteria in multiple databases, such difference between the AE using different algorithms are not significant.

Table 15. Average Error of synthesized patterns and exceptional patterns

S. No	Alpha	Gamma	Average Error
1	0.5	0.4	0.244432
2	0.5	0.5	0.169259
3	0.5	0.6	0.169259
4	0.5	0.7	0.600000
5	0.9	0.5	0.159999
6	0.5	0.9	0.850371
7	0.4	0.5	0.850371
8	0.5	0.5	0.130925
9	0.6	0.5	0.169259
10	0.7	0.5	0.175294

7. Conclusion

In this paper, a novel method has been projected for getting exact synthesized frequent items by clustering and exceptional patterns from the multiple data bases which are an imperative component of a multi database mining system. Several corporate decision of a multi branch company would depend on these two types of patterns in some situations to take decision effectively in the competitive world. The first type of the clustered frequent items which are used to provide the detailed information of association among the different type of items in a market place. The second type of exceptional patterns which are used to avoid the transportation cost of the company, major effort of manpower and non profitable items. These two tasks are the major role in the decision support systems in some special cases. When the items in a database are greatly associated, the proposed clustering and finding exceptional patterns techniques are working well. We have shown theoretically that the experimental results shows that the proposed clustering technique is efficient and effortless and produced exceptional and synthesized patterns are useful to take decision in higher level of the organization.

8. References

1. Ramkumar T, Srinivasan R. Modified algorithms for synthesizing high-frequency rules from different data sources. Knowl Inf Syst. 2008. 17:313–34.
2. Zhang S, Wu X, Zhang C. Multi-database mining. IEEE Computational Intelligence Bulletin. 2003; 2(1):5–13.
3. Zhang S, Zaki JM. Mining multiple data sources: Local pattern analysis. Data Min Knowl Discov. 2006; 12:121–5.
4. Adhikari A, Jain CL, Ramana S. Analysing effect of database grouping on multi-database mining. IEEE In-tell Inf Bull. 2011; 12:25–32.
5. Agrawal R, Srikant R. Fast algorithms for mining association rules. Proceedings of the International Conference on Very Large Data, Bases. 1994. p. 487–99.
6. Han J, Pei J, Yiwen Y. Mining frequent patterns without candidate generation. Proceedings of SIGMOD Conference on Management of Data. 2000. p. 1–12.
7. Hershberger SL, Fisher DG. Measures of Association, Encyclopedia of Statistics in Behavioral Science. John Wiley and Sons. 2005.
8. Frequent item set mining implementations repository. Available from: http://fimi.cs.helsinki.fi/src/
9. Han J, Kamber M. Data mining concepts and techniques. Morgan Kanufmann. 2000.
10. Adhikari A, Rao PR. Synthesizing heavy association rules from different real data sources. Pattern Recognition Letters. 2008; 29(1):59–71.
11. Adhikari, Rao PR, Adhikari A. Clustering items in different data sources induced by stability. Int Arab J Inform Tech. 2009; 4:394–402.
12. Zhang S, Wu X. Fundamentals of association rules in data mining and knowledge discovery. WIRESs Data Min Knowl Discov. 2011; 1:97–116.
13. Agrawal R, Imielinski T, Swami A. Mining association rules between sets of items in large databases. Proceedings of SIGMOD Conference on Management of Data. 1993. p. 207–16.
14. Agrawal R, Srikant R. Fast algorithms for mining association rules. Proceedings of the International Conference on Very Large Data, Bases. 1994. p. 487–99.
15. Hershberger SL, Fisher DG. Measures of Association, Encyclopedia of Statistics in Behavioral Science. John Wiley and Sons. 2005.
16. Adhikari A, Rao PR. Capturing association among items in a database. Data and Knowledge Engineering. 2008; 67(3):430–43.
17. Adhikari A. Synthesizing global exceptional patterns in different data sources. J Intell Syst. 2012; 21(3):293–323.
18. Zhang C, Liu M, Nie W, Zhang S. Identifying global exceptional patterns in multi-database mining. IEEE Computational Intelligence Bulletin. 2004; 3(1):19–24.
19. Wu X, Zhang C, Zhang S. Efficient mining of both positive and negative asso-ciation rules. ACM Trans Inform Syst. 2004; 22:381–405.
20. Wu X, Zhang C, Zhang S. Database classification for multi-database mining. Inform Syst. 2005; 30:71–88.
21. Ramkumar T, Srinivasan R. Modified algorithms for synthesizing high-frequency rules from different data sources. Knowl Inf Syst. 2008; 17:313–34.