ISSN 2347 - 3983
Volume 7, No.7 July 2019
International Journal of Emerging Trends in Engineering Research
Available Online at http://www.warse.org/IJETER/static/pdf/file/ijeter04772019.pdf
https://doi.org/10.30534/ijeter/2019/04772019

Mining Negative Associations between Regular and Frequent Patterns hidden in Static Databases

NVS Pavan Kumar[1], Dr.JKR Sastry[2], Dr. K Raja Sekhara Rao[3]
[1]Koneru Lakshmaiah Education Foundation, Vaddeswaram, Andhra Pradesh, India, nvspavankumar@kluniversity.in
[2]Koneru Lakshmaiah Education Foundation, Vaddeswaram, Andhra Pradesh, India, drsastry@kluniversity,in
3Usharama institute of Engineering and Technology, Andhra Pradesh, India, krr_it@yahoo.co.in

ABSTRACT

Most of the research in the field of data mining concentrated around finding the positive associations that exist among different kind of frequent patterns. Frequent patterns have no bearing on the time duration. The regularity of the patterns has a time stamping but no bearing on the frequency. The maximal property, in conjunction with regularity and frequency, leads to all important patterns. Negative associations among the mined patterns are quite important as they reveal significant contradictions which happen quite frequently in medical drugs and weather fluctuations.

The pattern mining also differs a lot based on the kind of data source used which may include static databases, distributed databases, incremental databases

In this paper, a method proposed that mines regular, frequent, and maximal patterns and the negative associations that exist among those patterns considering static databases

Key words: Data mining, Regular Patterns, Negative associations, Maximal patterns, frequent patterns, Static Database

1. INTRODUCTION

Transactional data can be stored either in a centralized database or distributed databases based on the way business is conducted centrally or at distributed locations. Data sometimes moves in streams, and the data in the streams generally takes more time to be placed into a database before it can be accessed. The nature of the data patterns existing in the database keeps changes as modifications to the databases are undertaken, especially when new data records added. Incrementing the existing data with new data alter the nature of the patterns existing in the database. In this chapter, the issue of finding the negative patterns / Associations from regular patterns considering fixed database presented.

A set of data items keeps occurring again and again and appears in different transactions. Frequency of occurrence of a set of items considering different transactions that happen over some time differs from the item set to item set. The set of data items together contained in a set is called a Pattern. There will be many such patterns in databases many in number. Not all the patterns are interesting

Frequent patterns (FP) are one of the fundamental knowledge discoveries recognized through transactional databases. A frequent itemset is the set of items that occur more than a user given minimum support threshold. Confidence and correlation are two more parameters included to increase the reliability of the frequent itemsets.
A frequent itemset (also called large item set [Chen et al. 1996]) is an item-set that meets the user-specified minimum support. Accordingly, we define an infrequent item set (or small item set) as an item set that does not meet the user-specified minimum support.

However, the frequent itemsets relate to occurrence frequency (number of occurrences) only but not how they spread in the distribution. Some of the patterns are sporadic and very rarely occur. Such kind of patterns is called outliers. They influence the mining process to move away from the real scenario. Many types of outliers exist in reality, which categorized as global, contextual, collective, and many more. The outliers eliminated before one arrives at frequent patterns.

Frequent Patterns play an important role in knowledge discovery databases. The patterns mined from a database classified as Basic patterns, Multi-level and Multi-dimensional patterns, and extended patterns. The Basic patterns include frequent patterns, Pattern of association rules, Closed Patterns, Maximum patterns, and generator patterns. Multi-level patterns are uniform, varied patterns or patterns that depend on the support of item-sets. Multi-dimensional patterns are high dimensional patterns. Some of the patterns classified as continuous patterns formed through the process of discretization or the application of statistical methods.

Extend patterns include approximate patterns, uncertain patterns, compressed patterns, rare or negative patterns and colossal patterns, high dimensional-colossal patterns.

Most attention made on finding the frequent items which focus on the frequency of occurrence of an item set. Not much focus used on finding the regular item sets. Regular itemsets are those that occur many times within a specific period. The periodicity could be absolute or just relative, which measured as the distance between the transactions that contain an item set. The regularity of the item set plays is one of the most important aspects that have a bearing on business decisions.

There could be an association between a set of patterns. The association between the patterns could be either positive or negative. In positively associated patterns, the same set of business rules equally apply, which means one pattern support the other. Most of the time business managers look for positively associated patterns

The association between the frequently occurring patterns will be more interesting and important as there is an issue of interestingness associated with such patterns. Also, the interestingness will be quite high if the issue of regularity considered. The positive association between regular and frequent will help in taking many important decisions.

Most of the attention is on mining frequent and positive patterns. Frequent patterns talk about the frequency of occurrence of the patterns. Regular patterns speak about both occurrence frequency and occurrence behavior. A period defined as the difference between two consecutive occurrences of the itemset. A regular itemset defined as the maximum of the periods of item set should be less than the user given regularity threshold. This property ensures the even distribution of the item set in a transactional database. Hence each item set must repeat within a given interval. Outliers will be automatically eliminated in this process as they behave differently from other objects.

Negative patterns are quite important even more than the positive patterns due to the kind of impact that it creates when such patterns exits. Negative patterns quite often are traced in medical fields, financial filed, and whether forecasting sector. In medical filed two drugs having different chemicals may contract each other. A rule that applies to a temperature zone may not apply to a cool zone. Thus the issues of finding the patterns that are regular, frequent, and those that have negative associations are most important which the focus of this research is. Modeling regular and negative patterns

Negative association rule mining is very much useful in this area. The term negative indicates the absence of an item. It also indicates the contradiction between two or more item sets. Recent days they are also identified by the name non-overlapping patterns.

Negative association rules are generated when the correlation between two items sets is negative; the confidence between the itemset is quite high, and this is true even if the support between the itemsets is not higher than the threshold value. Some of the itemsets could be regular but may not appear together. The itemsets of this nature are called regular itemsets.

Mining huge database created through several commercial applications for discovering frequent itemsets and association rules is complicated. The mining must be done to discover both negative and positive association rules required for decision making.

There are exceptional cases in negative patterns named as surprising patterns. Sometimes patterns behave surprisingly by deviating from the expected well-known fact. These patterns are exceptions of the association rules. Hence they are named as surprising patterns. Many exceptions as such exist. One such exception indicates that unexpected patterns and exceptional patterns can involve negative terms and therefore treated as a special case of negative rules.

The representation $\neg A$ is the negation of item A. The support of $\neg A$ obtained through subtracting support of the itemset A from 1.

Support (i1, ~i2, i3) = support (i1,i3) – sup(i1, i2, i3)

A rule of the form $A \Rightarrow B$ is considered to be a positive rule and other forms of the rules ($\neg A \Rightarrow \neg B$, B, $\neg A \Rightarrow B$, and, $A \Rightarrow \neg B$ are considered to be negative rules. The interestingness of a negative rule is the confidence of the rule expressed in terms of sup $(A \cup \neg B)$ / sup (A). The rules of the form $A \Rightarrow \neg B$ needs to be discovered which meets the minimum support and confidence provided by the user with the condition that A and B are disjoint sets.

—$supp(A) \geq ms$, $supp(B) \geq ms$ and $supp(A \cup B) < ms$;
—$supp(A \Rightarrow \neg B) = supp(A \cup \neg B)$;
—$conf(A \Rightarrow \neg B) = supp(A \cup \neg B)/supp(A) \geq mc$.

Rule $A \Rightarrow \neg B$ is referred to as an interesting negative rule. Association between the Frequent patterns (Positive and Negative)

Association rules are one of the precious outcomes of pattern mining. Several algorithms have developed to mine positive association rules. The term positive indicates the togetherness of the set of items wherever they appear in the transactional databases. These association rules developed from positive patterns or item set. If there exists two item sets A and B such that whenever an item set, A is purchased B is also purchased, then we define a positive association rule A =>B.

The strength of an association rule measured in terms of its supports, and confidence c. The support is the percentage of transactions in a database that contains all the elements, i.e., and A Ụ B. The confidence c is the percentage of transactions in a database containing A that also contain B. The formal definitions of these measures are,

$$\text{Support } (A \cup B) = \rho\,(A \cup B) / N$$
$$\text{Confidence } (A \Rightarrow B) = P\,(B \setminus A) = \text{support } (A \cup B) / \text{support } (A)$$

These expressions expressed in terms of probabilities. That is,

$$\text{Support } (A \Rightarrow B) = P\,(A \cup B)$$
$$\text{Confidence } (A \Rightarrow B) = P\,(B \setminus A)$$

Many associations exist such as ¬A⇒¬B, ¬A⇒B, A⇒¬B, and A⇒B, which makes the problem of finding the negative associations much complicated. Many problems arise in finding the frequent itemsets, infrequent itemsets, finding proper positive and negative association rules, the problem caused by single minimum support, and so on.

Many mining methods exist in the literature for different mining types of patterns. The Mining methods that are in existence as on date can be classified into basic mining methods that include candidate generation methods that have been many in number (Apriori, partitioning sampling etc.), pattern Growth methods (FP-growth, HMine, FP max, Closttt+ etc.), Vertical format methods (Eclat, CHARM etc.)

The Mining methods can also be classified based on the interestingness which includes interestingness (subjective Vs. Objective), Constraint-based mining, mining correlation rules, and exception rules

The mining methods also classified considering the way the database organized that includes a distributed database, incremental database, and streamed database.

Data mining methods greatly differ based on the type of database mined. Data mining methods can be classified based on the type of data mined. Various types of extended data that mined include sequential, time series, Structural (Tree, lattice, graph), spatial (Colocation), temporal (evolutionary and periodic0, Image Video and Multimedia and network patterns)

Mining methods classified that include pattern-based classification, clustering, semantic annotations, collaborative filtering, and Privacy-preserving.

There are some approaches presented in the literature for negative mining patterns. No method as such has been available that aims at mining regular, frequent negative patterns, which is the focus of this research. As said earlier, there are many types of databases exists, and the mining methods largely vary from type of database to type of database. In this chapter mining regular, frequent negative patterns through the use of vertical format, which is the basic mining method has been explored and presented.

2. PROBLEM DEFINITION

The problem is to find from the regular pattern, the patterns that have negative associations in the first instance and then move on to find the from frequent and regular patterns the patterns that have negative associations

3. RELATED WORK

Ming-Syan Chen [1] have presented a comprehensive survey on the availability of different mining methods and the purpose for which the mining methods used. A purposed based classification of the mining methods and the purpose for which the mining methods presented.

A basic method to mine transactional databases leads to mining too many patterns that will reflect into too many association rules that were not quite interesting to the end user. Ashok Savasere [2] Presented a mining method that combines positive associations with domain knowledge so that very few negative associations found which can be easily evaluated and presented.

Balaji Padmanabhan [3] Have presented that pattern mining generally leads to too many patterns and do not take into account the domain knowledge that the decision makers have. Decision makers have prior knowledge about the data in terms of precepts and beliefs. They have also presented a method that mines the unexpected patterns through the use of beliefs and perceptions and have experimented the method WEBlog files and proved that efficient mining could be done using the user's perception. Many mining methods existing in the literature have used the concept of candidate get generation approaches which uses Apriori like method. This approach has been proved to be time-consuming and costly, especially when many long patterns are involved.

Jiawe Han [4] have proposed a novel method which uses Frequent pattern tree (FP-tree), which is an extension of Prefix-structure. The FP-tree structure used for storing crucial information about the frequent patterns and the information used for pattern mining. The method used the concept of FP-growth for complete mining set of frequent patterns.

The method proposed by Jiawe Han [4] used three main techniques that include database compression, a pattern fragment growth method and divide and conquer method for decomposing mining tasks into the small number of tasks which can carry the mining considering the constraints attached to the small tasks. The three methods used by them reduce the search space dramatically.

Pattern mining can be undertaken either through horizontal or vertical mining approaches. Vertical mining approaches are found to be quite effective when compared to horizontal approaches. Fast frequency counting through intersecting operations on transactions IDs and pruning are irrelevant, which are the main advantages of vertical format methods. However, these methods suffer from lack of memory when the entries to be made into a vertical format table are too heavy. Mohammed J. Zaki [5] have presented a novel way of presenting the vertical data called Di-Set, which considers the differences between the transactions of a candidate patterns the very patterns itself. They have shown how Di-sets can drastically cut down the memory requirement for storing the vertical table entries,

In transactional databases, many patterns exist that can be used to generate both positive and negative association rules. A method has been proposed by XINDONG WU [6] that used for generating both negative and positive association rules. The negative associations between the patterns evaluated through checking the expressions like A⇒¬B, ¬A⇒B, and ¬A⇒¬B. The rules mined from a large database through constraining the patterns using the interesting patterns.

Some of the association rules that generated could be exceptional in the sense that the rules are less interested or have too high a confidence. Daly et al., [7] Have presented a method for exceptional mining rules evaluation. They have considered the relationships between the exceptional rules and negative association rules. Exception rules are generated based on the knowledge gained through negative association rules. They have also defined a new measure that can be used to evaluate the interestingness of the exceptional rules. The exceptional rules that meet the exceptional measures are the candidate's exceptional rules used for evaluation of the patterns and decision making.

Most of the methods proposed in the literature use the interestingness measures to prune the most wanted patterns for decision making. However, precisely defining the interestingness measure is quite complicated and sometimes must be found using the trial and error method. There is no such exact method used for determining the interesting measures. DR Thiruvady 2004 [8] presented a method that uses inputs provided by the user, in terms of several rules that the user requires and the kind of constraints/interestingness that must be satisfied. AN algorithm called GRD that discovers M-most interesting rules presented.

Correlations are statistical measures that find how good a set of data records are related to another set of records. Maria-Luiza, Antonie [9] have presented a method to find negative association rules, on the kind of existence of a correlation between two item sets. Negative rules between the item set extracted if the correlation between the Itemsets is negative and that the confidence of the items sets is quite high. The extracted negative association rules can have either consequents or antecedents (-Y¿ and $\neg Y$¿) even when the computed support value from the itemsets is less than the threshold value of the support. The algorithm presented by them generates all the negative and positive

The correlation between the item sets could be positive when the interesting measures hold good the interestingness measures that include support and confidence. If the support is less and confidence is high, then the negative correlation exists between the item sets, leading to a generation of negative association rules that negates both the antecedents and consequents.

The method proposed by Maria-Luiza, Antonie [9] generated all negative and positive associations rules out of the patterns which have a strong association between them. If no association rules generated, the threshold Value of correlation will have to be lowered, thereby reducing the strength of correlations between the items set.

A survey has been presented by Chris Cornelis [10] citing several algorithms that mine both negative and positive association rules and have described several situations wherein the algorithms presented in the literature could not satisfy certain situations. They have classified and cataloged several mining algorithms based on some parameters and could figure out the drawbacks of each of the algorithms. They have also presented a modified mining algorithm based on Apriori approach that can find both negative associations with interesting attached through confidence framework. They have used upward closure property that conforms to the support based interestingness of negative associations under validity definitions.

Usually, the interestingness parameter "Support" defined for entering dataset. The data records recognized as a hierarchy of records having the occurrence of a set of records at each level having a specific support value. Several support values defined at each level. A model is proposed by [Xiangjun Dong [11] called MLMS (Multi-Level minimum support) that considers defining minimum support value at each of the levels of the records. MLMS is used to discover both frequent and infrequent itemsets. They have considered both correlation and confidence interesting measures and proposed yet another interesting measure to mine both frequent and infrequent itemsets. An algorithm called PNAR-MLMS has proposed that can be used to generate both positively and negatively associated patterns from frequent and infrequent itemsets generated through MNMS model.

[Xiangjun Dong 2007-1] [12] have also developed PNAR based Classifiers using which the association rules classified into some known categories. The classifiers then can be used to find whether a pattern leads to negative or positive association. Discovering K-Most intersecting rule requires the minimum support value, which is the threshold value. It is rather difficult to define minimum thresh hold value as the user has no real idea about the support value. Rather the users can define interestingness and the number of rules that the users expect from the mining system.

Another method called GRD, which does not require minimum support value, also presented in the literature. It mealy requires the user to define the measure of interestingness and the number of rules in which the user is interested in. Xiangjun Dong [13] have extended the GRD method, which used as a form for mining positive and negative rules.

Both positive and negative association rules mined through transactions. Negative association rules explain how one pattern negates some other patterns. Many applications exist that needs the mining of the negative association rules, especially the negative association rules, will help in carrying the market-basket analysis.

The negative association rules can be used to develop classifiers as well using which classification models built. Mining negative association rules require the exploration of large data space. Many of the algorithms proposed in the literature are not in use due to the usage of large databases.

Xiangjun Dong et al., [14] Have extended the support confidence framework through the addition of correlation coefficient threshold that keeps sliding as the data accessing keeps moving. Essentially they have used correlation coefficients that can be calculated considering different patterns. The patterns that are positively and negatively correlated found form antecedents and consequents.

Most of the work is focussed on frequent items till 2009 and no focus made till such time on regular items which are those that occur in regular intervals. Regular items are the most important than the frequent items, the occurrence of which has no time limitation. **Tanbeer et al., [15]** are the first group of authors who have focussed on regular itemsets. They have proposed a “Regular pattern tree,” which is a tree structure to discover regular patterns. The algorithm scans the database twice. In the first set regularity and support values of the item, sets are determined, and in the second scan, a regular Pattern tree constructed. The process adopted by them is similar to cyclic and periodic patterns.

In many transactional databases, data hidden in sequence as a structure. Bio-Technology based sequences hidden in related medical databases. Mining sequential patterns reveal many interesting facts that when evaluated yield important decisions. Generally, the sequential patterns are mined using defined minimum support threshold defined by the users. Use of the minimum support threshold assumes that all frequent sequences have the same frequency, which is not the case in the real world. If the frequencies of the pattern sequences vary even though they meet the minimum threshold value, then a rare item problem arises**.**

Mining negative associations are as important as mining positive associations among the frequent patterns. Idheba Mohamad Ali [17] Have presented new models used for mining interesting negative and positive associations among transactional data records. They have considered the merging of two algorithms that include mining interesting negative and positive association rules (PNAR) and mining interesting multiple level supports algorithm (IMAMS). The algorithm proposed by them helps mining positive and negative association rules from interesting frequent and in-frequent item sets using multiple support values.

NVS Pavan [18] [19] have presented a method of finding both positive and negative association considering the regularity of the item set using vertical Table mining method.

Many mining methods presented in the literature [20][21][22][23][24][25][26][27][28][29][30] not covering the issue of regularity and negativity of the associations between regular and frequent patterns.

4. COMPARATIVE ANALYSIS

Comparison of existing algorithms has been made considering various aspects considered for generating negative associations considering regular and frequent itemsets to assess the adequacy of those algorithms. Table 1 shows the comparison, from the table it can be seen that none of the existing algorithms are dealing with the most important aspects of the negative associations that include regularity, positive/negative associations, frequency and interestingness measures.

5. INVESTIGATION AND FINDINGS

5.1 Algorithm

1. Read the support value that dictates the threshold value of the frequency of the patterns and also the regularity as defined by the user
2. Read the data in flat file / DBMS Table into an Array as shown in Table 1
3. Convert the data in table 1 into the vertical format as shown in Table 2
4. Prune the Initial Irregular Items and non-frequent items
5. Repeat the following process Consider the current Item

   Select next item and prune it if it is not regular or frequent and go to the next item (self-Loop).

   If the next item is regular and frequent, get the intersection of the transactions of the current item and next item

   If the intersection is null, then enter the current and next items into a negative set array

   If the intersection is not null, then get the common elements and see if the count of elements is > regularity threshold decided by the user

   If the common elements satisfy the regularity and frequency constraint, then add the common elements into a vertical table as a new row as they are regular and frequent at the end of the vertical table.

   If the common elements do not satisfy the regularity or frequency constraint, then ignore them

   If all the elements in the vertical table are exhausted, then convert the next item which is next to the current item as the current item and then LOOP

   If all the elements in the vertical table are not exhausted, then move to the next item and LOOP

   When no item left in the vertical table, then the process terminates

6. Find the Patterns from the negative pattern list

5.2 Experimentation using Sample Data derived out of IBM supplied database

The proposed algorithm applied on a sample of data derived out of IBM supplied data set, and the following results are applied

Step-1

Transaction IDS added to the IBM supplied data. Sample first 21 records of IBM supplied data shown in Table 2. Here frequency is the count of transactions in which the item set appears

Table 2: Sample data with Transaction ID

TrId	Item Set
1	I1 I2 I3 I4 I5 I9 I10 I14
2	I4 I5 I6 I10 I15
3	I2 I3 I7 I13 I14 I15
4	I5 I8 I10 I11 I15
5	I1 I3 I5 I6 I9
6	I4 I5 I6 I15
7	I2 I3 I7 I11 I12 I13
8	I5 I8 I11 I12 I14 I15
9	I1 I3 I5 I8 I9
10	I4 I5 I6 I10 I15
11	I2 I3 I7 I8 I13 I14 I15
12	I5 I8 I11 I15
13	I1 I3 I5 I9 I11
14	I4 I5 I6 I14 I15
15	I2 I3 I6 I7 I12 I13
16	I5 I8 I11 I12 I14 I15
17	I1 I3 I5 I6 I9 I10
18	I4 I5 I6 I12 I14 I15
19	I2 I3 I4 I7 I13
20	I5 I8 I11 I12 I15
21	I1 I3 I5 I9 I14

Step-2

Convert the Transaction filled IBM data to Vertical format. The vertical format data have shown in Table 3. In the vertical data format, for each of the Item, in the data repository, the transactions that contain the Items are found and mapped. The frequency of an item is the count of transactions in which the item appears.

Table 3: Transaction Data in vertical format

Item Code	TrId
I1	1 5 9 13 17 21
I2	1 3 7 11 15 19
I3	1 3 5 7 9 11 13 15 17 19 21
I4	1 2 6 10 14 18 19
I5	1 2 4 5 6 8 9 10 12 13 14 16 17 18 20 21
I6	2 5 6 10 14 15 17 18
I7	3 7 11 15 19
I8	4 8 9 11 12 16 20
I9	1 5 9 13 17 21
I10	1 2 4 10 17
I11	4 7 8 12 13 16 20
I12	7 8 15 16 18 20
I13	3 7 11 15 19
I14	1 3 8 11 14 16 18 21
I15	2 3 4 6 8 10 12 14 16 18 20

Step-3

Find the first regular item by pruning all the previous items whose regularity is > User given Maximum Regularity threshold ($ý_{min\_reg}$) and also the item that satisfies the minimum support value. Here regularity implies the relative occurrence of the Item, computed as the distance between two successive transactions. Consider ($ý_{min\_reg}$) = 4 and the minimum frequency be 3 for sample data. The First regular and frequent Item is called Previous-Item. The list of items that will be left over in the vertical format table shown in Table 3

Step-4

Consider each item starting from Previous-item and repeat the following procedure.

1. Consider the next item and let that be current-item
2. Find if the current-item is regular and frequent. If the current item is not regular or frequent prune it.
3. If the current- item is regular and frequent, find Intersection of the transactions of the current item with the previous-item.
4. If the intersection is null, then add the Item set into a negative item-set list.
5. If the intersection is not null, find the regularity considering the common elements.
6. If the regularity is < ($ý_{min\_reg}$), then add the previous item and the current item set along with its related transaction as an additional record to the vertical database since they are regularly and frequently associated.
7. If the next item is not the lost entry in the vertical table, Make Current Item as the next Item and loop.

8. If the next item is the last item in the Vertical table then Previous Item = Previous Item +1 and then Loop.

Step-5

Find the Maximum Item sets from negatively associated patterns

After this step completed, Pruned Items in the vertical table until the first regular and frequent Item traced is shown in Table 4, the pruned items shown in Table 5, and the negatively associated items in Table 6, and positively associated items in Table 7, and the Maximum Item sets are in Table 8.

Table 5: List of a Pruned item list

Item Code	TrId	Regularity (Periods)	Maximum Regularity of the Item
I10	1 2 4 10 17	1 1 2 6 7 4	7
I12	7 8 15 16 18 20	7 1 7 1 2 2 1	7
I14	1 3 8 11 14 16 18 21	1 2 5 3 3 2 2 3	5

Table 6: Negatively Associated Item set

Item Set-1	Item-set 2	Item Set-1	Item-set 2
1	**7**	**7**	**5,9**
1	**13**	**7**	**1,5,9**
1	**15**	**8**	**4,6**
1	**7,13**	**9**	**13**
4	**8**	**9**	**15**
4	**11**	**9**	**7,13**
4	**8, 11**	**11**	**4,6**
5	**7**	**13**	**1,5**
5	**7,13**	**13**	**1,9**
6	**8**	**13**	**1,5,9**
6	**8**	**15**	**1,9**
6	**11**	**1,5**	**7,13**
7	**9**	**4,6**	**8,11**
7	**1,5**	**5,9**	**7,13**
7	**1,9**	**1,5,9**	**7,13**

5.3 Pseudo Code

Notations

Let T = {t_1, t_2, t_3 … t_m} be an m transaction constitutes the database DB. Each transaction t € T is of the form (TrId, S) where TrId stands for Transaction ID which is an integer and S is a set of items (pattern). All the items i_k in transaction *t* belongs to the set of items of the database denoted by I={i_1, i_2, i_3,…. i_n}. Let the number of item in the database are *n*, and the total number of transactions present in the database are m=|DB|. Hence 0< t_k<=m for all t_k. The pattern S= {i_p, i_q , … i_r} is regular when the maximum regularity of this pattern is less than the user-defined minimum regularity threshold $\alpha_{min\_reg.}$ Two patterns S1 and S2 are negative regular Itemset if they are regular and $T_{S1} \cap T_{S2}$ ={Φ}.

```
Algorithm NPRISM ( )
{
        // To find first regular item

 for k=1 to n

{       if ( regularity( Ik, TrId^Ik, λ max_reg, λ min-freq) = = FALSE)
            delete Ik;;
          else
            break;
  }

// To find remaining regular items

for  j= k+1 to n
{
    if  ( regularity( Ij, TrId^Ij, λ max_reg, λ min-freq)= = FALSE)
      delete Ij;
    else
    {
      PRISM(Ik,Ij);
      NRI(ik,ij);
    }
}

Boolean Regularity (Ik, TrId^Ik, λ max_reg, λ min-freq)
{

// TrId^Ik first and TrId^Ik last are the first and last transactions of Ik

        Reg_First =TrId^Ik first-0;

        if ( (Reg_First > λ max_reg) OR   (Freq_First > λ min-freq)
return FALSE;
            Reg_Last = m-TrId^Ik last;

        if ((Reg_Last > λ max_reg) OR (Freq-Last > λ min-freq)  return
FALSE;
            Reg_gen={};

        for p= first+1 to last
        {
                Reg_gen = Reg_gen ∪ {TrId^Ik p - TrId^Ik p-1};
        }
        if (MAX{Reg_gen}> λ max_reg)        return FALSE;

        return TRUE;
}
```

Algorithm PRISM(I_k,I_j)
{
$I_{kj}= I_k \cup I_j$;
$TrId^{Ikj} = TrId^{Ik} \cup TrId^{Ij}$;
if (regularity(I_{kj}, $TrId^{Ikj}$, λ_{max_reg})= = FALSE)
delete I_{kj} ;

else

{
VDB= VDB $\cup$ { I_{kj} , $TrId^{Ikj}$ };
n=n+1;
}
}

Algorithm NRI (I_k,I_j)

{
if (I_{kj} < VDB) return;
if (TrIdIk(TrIdIj) = = {})
NegItemSet=NegItemSet $\cup I_{kj}$;
else
return ;
}

6. DATA ANALYSIS AND INTERPRETATIONS

The IBM supplied data has been analysed for different sizes of the samples drawn in terms of 40,000, 60,000 and 80,000 records. The data is analysed with different regularity percentages. The number of negative frequent regular patterns, for different maximum regularity and minimum frequency is shown in Table 9

Table 9: Analysis of Negative Frequent Regular Itemsets with TD 100K Transaction data

Total Transactions	%Max Regularity	%Support Count	Number of Negative Frequent Regular Itemset
40000	3.00	2.000	6
	3.00	1.750	42
	3.00	1.625	154
	3.00	1.500	461
40000	2.50	1.750	41
	2.50	1.625	154
	2.50	1.500	352
	2.50	1.125	981
40000	2.00	1.750	35
	2.00	1.625	118
	2.00	1.500	181
40000	1.50	1.750	2
	1.50	1.625	3
	1.50	1.500	3
60000	1.65	1.650	13
	1.65	1,250	41
	1.65	1.000	150
	1.65	1.000	352
60000	1.35	1.650	35
	1.35	1.350	118
	1.35	1.000	181
80000	1.00	0.875	35
	1.00	0.815	118
	1.00	0.750	181

Figure 1 refers to the number of negative frequent regular and negative regular items in relation to % regularity. The numbers of frequent regular items sets are narrowing down compared to regular frequent sets as the % of regularity and support count increases.

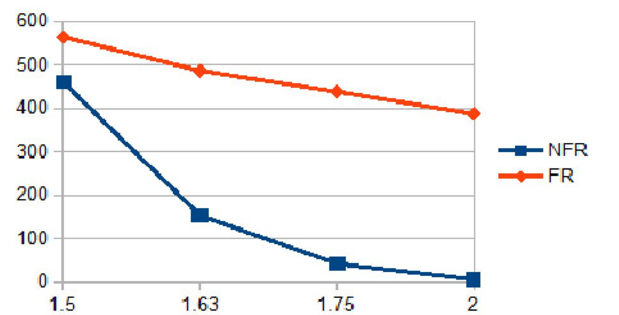

Figure 1: Negative Regular and Negative Frequent Regular with 40000 transactions at different Regularity and Frequency values.

Figure 2, Figure 3, Figure 4 shows the relationship between the negative frequent regular item sets and frequency plotted on the x-axis for different regularity %. It could be noted from the figures that as the frequency increases the number of item sets decreases.

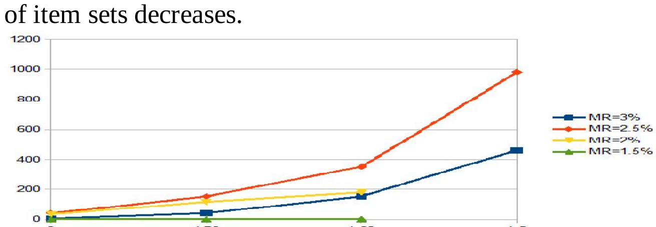

Figure 2: Negative Frequent Regular Itemset with 60000Transactions at different Regularity and Frequency values

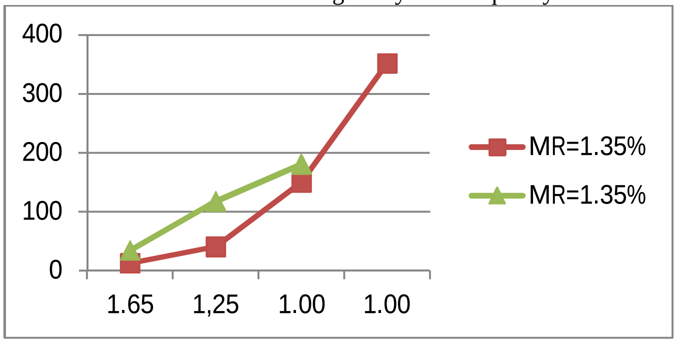

Figure 3: Negative Frequent Regular Itemset with 60000Transactions at different Regularity and Frequency values

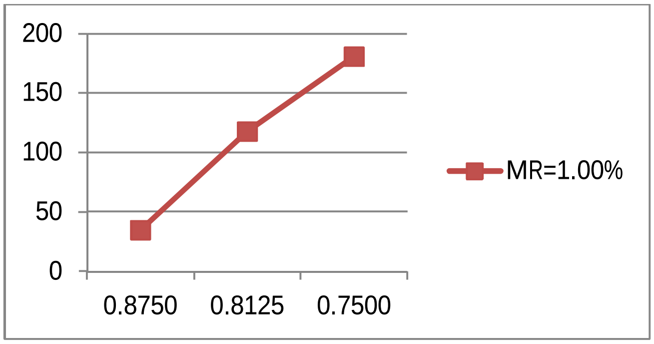

Table 4: Negative Frequent Regular Itemset with 80000 Transactions at different Regularity and Frequency values

7. CONCLUSION

Many types of databases exist which differ in the kind of data stored. Transactional databases stored transactional data. Transactional data could be stored either as flat files or relational data. Patterns are the knowledge hidden in any database. A set of data items that frequently occurs in different records in a database called patterns. Knowledge of patterns is very important as one can take excellent decisions by analyzing and visualizing the patterns

A set of patterns can be either frequent or regular — frequency of patterns computed considering the entire database. Regularity is a kind of distance between to like patterns. The frequency of the patterns within Regularity is also equally important.

Most concentration as on date is finding the frequent patterns which are positively associated. Regularity is also most important in addition to the frequency. Many negatively associated patterns exist in nature that needs to consider the ill effects of patterns on the business. A lot of importance for the negatively associated patterns exists in the Medical and atmospheric studies fields. Regularly negative, regularly frequent and negative, regularly frequent maximal and negative patterns are to be mined and analyzed so that analysis made and proper decisions are taken based on those patterns. It noted that the frequent patterns might not regular and Vice versa.

REFERENCES

1. Ming-Syan Chen, Jiawei Han, Philip S. Yu, Datamining an overview from database perspective IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL 8, NO 6, DECEMBER 1996 https://doi.org/10.1109/69.553155
2. Ashok SAVASERE, A., OMIECINSKI, E., AND NAVATHE, S. 1998. Mining for strong negative associations in a large database of customer transactions. In Proceedings of the Fourteenth International Conference on Data Engineering. IEEE Computer Society, Orlando, Florida, 494–502.
3. Balagi PADMANABHAN, B. AND TUZHILIN, A. 1998. A belief-driven method for discovering unexpected patterns. In Proceedings of the Fourth International Conference on Knowledge Discovery and Data Mining (KDD-98). AAAI, Newport Beach, California, USA, 94–100.
4. Jaiwe Han, Yin, Y. Yin, “Mining Frequent Patterns without candidate generation”, In Proc. ACM SIGMOD International Conference on Management of Data, PP. 1-12 (2000). https://doi.org/10.1145/335191.335372
5. Mohammed J. Zaki, Fast Vertical Mining Using Diffsets, *SIGKDD* 2003, ACM Copyright 2003
6. XINDONG Wu, C. Zhang, and S. Zhang, "Efficient Mining of both Positive and Negative Association Rules," ACM Transactions on Information Systems, 22, 2004, pp. 381-405. https://doi.org/10.1145/1010614.1010616
7. O. Daly and D. Taniar, ”Exception Rules Mining Based On Negative Association Rules,” Lecture Notes in Computer Science, Vol. 3046, 2004, pp 543–552 https://doi.org/10.1007/978-3-540-24768-5_58
8. D.R. Thiruvady and G.I. Webb, ”Mining Negative Association Rules Using GRD,” Proc. Pacific-Asia Conf. on Advances in Knowledge Discovery and Data Mining, 2004, pp 161–165. https://doi.org/10.1007/978-3-540-24775-3_20
9. Maria Luiza Antonie, O. Zaiane, "Mining Positive and Negative Association Rules An Approach for Confined Rules," Proceedings of 8th European Conference on Principles and Practice of Knowledge Discovery in Databases(PKDD04), LNCS 3202, Springer-Verlag Berlin Heidelberg, Pisa, Italy, 2004, pp.27-38. https://doi.org/10.1007/978-3-540-30116-5_6
10. Chris Cornelis, Ghent, Peng Yan, Xing Zhang, Guoqing Chen, Mining Positive and Negative Association Rules from Large Databases, IEEE 2006 https://doi.org/10.1109/ICCIS.2006.252251
11. Xiangjun Dong, Sun, F., Han, X., Hou, R.: Study of Positive and Negative Association Rules Based on Multi-confidence and Chi-Squared Test. In: Li, X.,

Zaïane, O.R., Li, Z. (eds.) ADMA 2006. LNCS (LNAI), vol. 4093, pp. 100–109. Springer, Heidelberg (2006) https://doi.org/10.1007/11811305_10

12. Xiangjun Dong, Z Niu, X Shi, X Zhang, D. Zhu, "Mining Both Positive and Negative Association Rules from Frequent and infrequent Itemsets," Proceedings of the Third International Conference on Advanced Data Mining and Applications (ADMA 2007), Harbin, 519 China, August 6-8, 2007. Lecture Notes in Computer Science 4632, Springer 2007, pp. 122-133. https://doi.org/10.1007/978-3-540-73871-8_13
13. Xiangjun Dong, Zheng, Z., Niu, Z., Jia, Q.: Mining Infrequent Item sets based on Multiple Level Minimum Supports. In: ICICIC07. Proceedings of the Second International Conference on Innovative Computing, Information and Control, Kumamoto, Japan, September 2007
14. Xiangjun Dong, Zhendong Niu, Donghua Zhu, Zhiyun Zheng[4], and Qiuting Jia, Mining Interesting Infrequent, and Frequent Itemsets Based on MLMS Model, C. Tang et al. (Eds.): ADMA 2008, LNAI 5139, pp. 444–451, 2008
15. TANBEER Syed Khairuzzaman, Chowdary Farhan Ahmed, Byeong-Soo Jeong, Young-Koo Lee, Mining regular patterns in transactional databases, IEICE Transactions, INF, & SYST., Vol. E91-D, Issue. 11, 2008
16. Weimin Ouyang, Qinhua Huang, Mining Positive and Negative Sequential Patterns with Multiple Minimum Supports in Large Transaction Databases, 2010 Second WRI Global Congress on Intelligent Systems https://doi.org/10.1109/GCIS.2010.213
17. Idheba Mohamad Ali O. Swesi, Azuraliza Abu Bakar, Anis Suhailis Abdul Kadir, Mining Positive and Negative Association Rules from Interesting Frequent and Infrequent Itemsets, 2012 9th International Conference on Fuzzy Systems and Knowledge Discovery (FSKD 2012)
18. N.V.S. Pavan Kumar, K Rajasekhara Rao, Mining Positive and Negative Regular Item-Sets using Vertical Databases, DOI 10.5013/IJSSST.a.17.32.33 33.1 ISSN: 1473-804x online, 1473-8031
19. N.V.S.Pavan Kumar, L. Jaga Jeevan Rao, G. Vijay Kumar, **A Study on Positive and Negative Association rule mining,** International Journal of Engineering Research & Technology (IJERT), Vol. 1 Issue 6, August – 2012.
20. Changala, R., Rajeswara Rao, D., Evaluation and analysis of discovered patterns using pattern classification methods in text mining, ARPN Journal of Engineering and Applied Sciences, 13(11), pp. 3706-3717,2018
21. Kolli, S., Sreedevi, M., Prototype analysis of different data mining classification and clustering approaches, ARPN Journal of Engineering and Applied Sciences,13(9), pp. 3129-3135,2018
22. Deshpande, L., Rao, M.N., Concept drift identification using classifier ensemble approach, International Journal of Electrical and Computer Engineering,8(1), pp. 19-25,2018 https://doi.org/10.11591/ijece.v8i1.pp19-25
23. Changala, R., Rajeswara Rao, D. T., A survey on the development of pattern evolving model for discovery of patterns in text mining using data mining techniques, Journal of Theoretical and Applied Information Technology,95(16), pp. 3974-3981,2017
24. Waghere, S.S., Rajarajeswari, P.., Parallel frequent dataset mining and feature subset selection for high dimensional data on Hadoop using map-reduce, International Journal of Applied Engineering Research, 12(18), pp. 7783-7789,2017
25. Vijay Kumar, G., Krishna Chaitanya, T., Pratap, M., Mining popular patterns from the multidimensional database, Indian Journal of Science and Technology, 9(17),93106 https://doi.org/10.17485/ijst/2016/v9i17/93106
26. Gangadhar, M.N.S., Sreedevi, M., Regular pattern mining on dynamic databases using vertical formate on given user regularity threshold, Journal of Theoretical and Applied Information Technology, 86(3), pp. 360-364, 2016
27. Greeshma, L., Pradeepini, G., Input split frequent pattern tree using MapReduce paradigm in Hadoop, Journal of Theoretical and Applied Information Technology, 84(2), pp. 260-271
28. Greeshma, L., Pradeepini, G., Mining Maximal Efficient Closed Itemsets Without Any Redundancy, Advances in Intelligent Systems and Computing, 433, pp. 339-347 https://doi.org/10.1007/978-81-322-2755-7_36
29. Real-time Streaming Data Analysis using Spark, International Journal of Emerging Trends in Engineering Research, Volume 6, No.1, 2018, pp. 1-5 https://doi.org/10.30534/ijeter/2018/01612018
30. Career Guidance through TIC-TAC-TOE Game, International Journal of Emerging Trends in Engineering Research, Volume 7, No.6, 2019, pp. 25-31 https://doi.org/10.30534/ijeter/2019/01762019

Table 1: Comparative Analysis of Pattern finding Algorithms

Algorithm Serial Number	**Main Author**	**Interestingness measures**					**Occurrence Behaviour**					**Type of Associations**		**Extension to Mining technique**	**Use of Domain Knowledge**
		Support	**Confidence**	**Correlation**	**Multi support**	**Multi Correlation**	**Regularity**	**Irregularity/Rare**	**Frequent**	**Maximal**	**Unexpected**	**Positive Associations**	**Negative Associations**		
1	Ashok Savasere	√										√			√
2	**Balaji Padmanabhan**	√									√	√			√
3	Jiawe Han 2000	√							√			√		FP Tree	
4	J. Zaki	√							√			√		DI-SET	
5	XINDONG WU	√							√			√	√		
6	Daly	√							√				√	Exception rule Mining	
7	DR Thiruvady	√											√		√
8	Maria-Luiza, Antonie			√					√			√	√		
9	Xiangjun Dong			√	√							√	√		
10	**Tanbeer**						√								
11	**Weimin Ouyang**				√									Sequential Mining	
12	Idheba Mohamad Ali				√		√					√	√		
13	Pavan NVS	√					√	√	√	√	√	√	√	Veridical Tab	

Table 4: Pruning the Items in the vertical table until the first regular and frequent Item traced

Item Code	TrId	Regularity (Periods)	Maximum Regularity
I1	1 5 9 13 17 21	1 4 4 4 4	4
I2	1 3 7 11 15 19	1 4 4 4 4 2	4
I3	1 3 5 7 9 11 13 15 17 19 21	1 2 2 2 2 2 2 2 2 2 2	2
I4	1 2 6 10 14 18 19	1 1 4 4 4 4 1 2	4
I5	1 2 4 5 6 8 9 10 12 13 14 17	1 1 2 1 1 2 1 1 2 1 1 2 1 1 2 1 3	3
I6	2 5 6 10 14 15 17 18	2 3 1 4 4 1 2 1 3	4
I7	3 7 11 15 19	3 4 4 4 4 2	4
I8	4 8 9 11 12 16 20	4 4 1 2 1 4 4 1	4
I9	1 5 9 13 17 21	1 4 4 4 4 4	4
I11	4 7 8 12 13 16 20	4 3 1 4 1 3 4 1	4
I13	3 7 11 15 19	3 4 4 4 4 2	4
I15	2 3 4 6 8 10 12 14 16 18 20	2 1 1 2 2 2 2 2 2 2 1	2

Table 7: Positively Associated Item set

Itemset	**Trids**	**Periods**	**Max Regularity**
1	**1 5 9 13 17 21**	**4 4 4 4 4**	**4**
2	**1 3 7 11 15 19**	**2 4 4 4 4**	**4**
3	**1 3 5 7 9 11 13 15 17 19 21**	**2 2 2 2 2 2 2 2 2 2**	**2**
4	**1 2 6 10 14 18 19**	**1 4 4 4 4 1**	**4**
5	**1 2 4 5 6 8 9 10 12 13 14 16 17 18 20 21**	**1 2 1 1 2 1 1 2 1 1 2 1 1 2 1**	**2**
6	**2 5 6 10 14 15 17 18**	**3 1 4 4 1 2 1**	**4**
7	**3 7 11 15 19**	**4 4 4 4**	**4**
8	**4 8 9 11 12 16 20**	**4 1 2 1 4 4**	**4**
9	**1 5 9 13 17 21**	**4 4 4 4 4**	**4**
11	**4 7 8 12 13 16 20**	**3 1 4 1 3 4**	**4**
13	**3 7 11 15 19**	**4 4 4 4**	**4**
15	**2 3 4 6 8 10 12 14 16 18 20**	**1 1 2 2 2 2 2 2 2 2**	**2**
1,3	**1 5 9 13 17 21**	**4 4 4 4 4**	**4**
1,5	**1 5 9 13 17 21**	**4 4 4 4 4**	**4**
1,9	**1 5 9 13 17 21**	**4 4 4 4 4**	**4**

Itemset	Trids	Periods	Max Regularity
2,3	1 3 7 11 15 19	2 4 4 4 4	4
2,7	3 7 11 15 19	4 4 4 4	4
2,13	3 7 11 15 19	4 4 4 4	4
3,5	1 5 9 13 17 21	4 4 4 4 4	4
3,7	3 7 11 15 19	4 4 4 4	4
3,9	1 5 9 13 17 21	4 4 4 4 4	4
3,13	3 7 11 15 19	4 4 4 4	4
4,5	1 2 6 10 14 18	1 4 4 4 4	4
4,6	2 6 10 14 18	4 4 4 4	4
4,15	2 6 10 14 18	4 4 4 4	4
5,6	2 5 6 10 14 17 18	3 1 4 4 3 1	4
5,8	4 8 9 12 16 20	4 1 3 4 4	4
5,9	1 5 9 13 17 21	4 4 4 4 4	4
5,11	4 8 12 13 16 20	4 4 1 3 4	4
5,15	2 4 6 8 10 12 14 16 18 20	2 2 2 2 2 2 2 2 2	2
6,15	2 6 10 14 18	4 4 4 4	4
7,13	3 7 11 15 19	4 4 4 4	4
8,11	4 8 12 16 20	4 4 4 4	4
8,15	4 8 12 16 20	4 4 4 4	4
1,3,5	1 5 9 13 17 21	4 4 4 4 4	4
1,3,9	1 5 9 13 17 21	4 4 4 4 4	4
2,3,7	3 7 11 15 19	4 4 4 4	4
2,3,13	3 7 11 15 19	4 4 4 4	4
11,15	4 8 12 16 20	4 4 4 4	4
1,3,5	1 5 9 13 17 21	4 4 4 4 4	4
1,5,9	1 5 9 13 17 21	4 4 4 4 4	4
3,5,9	1 5 9 13 17 21	4 4 4 4 4	4
4,5,6	2 6 10 14 18	4 4 4 4	4
4,5,15	2 6 10 14 18	4 4 4 4	4
4,6,15	2 6 10 14 18	4 4 4 4	4
5,6,15	2 6 10 14 18	4 4 4 4	4
2,3,7	3 7 11 15 19	4 4 4 4	4
2,7,13	3 7 11 15 19	4 4 4 4	4
3,7,13	3 7 11 15 19	4 4 4 4	4
5,8,11	4 8 12 16 20	4 4 4 4	4
5,8,15	4 8 12 16 20	4 4 4 4	4
1,3,9	1 5 9 13 17 21	4 4 4 4 4	4
1,5,9	1 5 9 13 17 21	4 4 4 4 4	4
3,5,9	1 5 9 13 17 21	4 4 4 4 4	4
5,8,11	4 8 12 16 20	4 4 4 4	4
5,11,15	4 8 12 16 20	4 4 4 4	4
8,11,15	4 8 12 16 20	4 4 4 4	4
2,3,13	3 7 11 15 19	4 4 4 4	4

Itemset	Trids	Periods	Max Regularity
2,7,13	3 7 11 15 19	4 4 4 4	4
3,7,13	3 7 11 15 19	4 4 4 4	4
4,5,15	2 6 10 14 18	4 4 4 4	4
4,6,15	2 6 10 14 18	4 4 4 4	4
5,6,15	2 6 10 14 18	4 4 4 4	4
5,8,15	4 8 12 16 20	4 4 4 4	4
5,11,15	4 8 12 16 20	4 4 4 4	4
8,11,15	4 8 12 16 20	4 4 4 4	4
4,5,6,15	2 6 10 14 18	4 4 4 4	4
2,3,7,13	3 7 11 15 19	4 4 4 4	4
5,8,11,15	4 8 12 16 20	4 4 4 4	4
2,3,7,13	3 7 11 15 19	4 4 4 4	4
4,5,6,15	2 6 10 14 18	4 4 4 4	4
5,8,11,15	4 8 12 16 20	4 4 4 4	4
1,3,5,9	1 5 9 13 17 21	4 4 4 4 4	4

Table 8 Maximal Itemset

Item Set-1	Item-set 2
7	1,5,9
13	1,5,9
1,5	7,13
1,9	7,13
4,6	8,11
5,9	7,13
1,5,9	7,13