

Article

Not peer-reviewed version

Mining Negative Associations From Medical Databases Considering Frequent, Regular, Closed and Maximal Patterns

[Sastry Kodanda Rama Jammalamadaka](#) * and [Reja Rao Budaraju](#)

Posted Date: 14 November 2023

doi: 10.20944/preprints202311.0905.v1

Keywords: data mining; databases; closed item sets; maximal item sets; regular patterns; frequent patterns; negative associations; maximal patterns; frequent patterns; static



Preprints.org is a free multidiscipline platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Article

Mining Negative Associations from Medical Databases Considering Frequent, Regular, Closed and Maximal Patterns

Sastry Kodanda Rama Jammalamadaka ^{2,*} and Raja Rao Budaraju ¹

¹ Scholar, Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Guntur-522302, Andhra Pradesh, India;

² Professor, Department of Electronics and Computer Science, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Guntur-522302, Andhra Pradesh, India; ; drsastry@kluniversity.in

* Correspondence: drsastry@kluniversity.in

Abstract: Most of the research in data mining concentrated on finding the positive associations that exist in frequent patterns. Frequent patterns have no bearing on the time duration and, therefore, lose interestingness. The Regularity of an itemset is related to the time at which the transaction occurred and the time distance between the transactions. Frequent Regular item sets can be generated considering the Frequency and the Regularity of the occurrence of the Item Sets. The frequent and regular Item sets will be so huge that they cannot be handled in real time. The count of Frequent and regular item sets gets reduced drastically when the properties “closed” and “Maximality” are applied. Finding negative associations among such patterns are important in the field of medicine. Many drugs administered to cure some diseases have counter reactions leading to many other diseases. Negative associations among the mined patterns are important as they reveal significant contradictions frequently among medical drugs administered by medical practitioners to cure some diseases. This paper proposes a method that mines medical databases to find regular, frequent patterns applied with closed and maximal properties and find negative associations among those patterns. The proposed algorithm is accurate to 98% and performs 9% more efficiently than the other algorithms presented in the literature.

Keywords: data mining; databases; closed item sets; maximal item sets; regular patterns; frequent patterns; negative associations; maximal patterns; frequent patterns; Static

1. Introduction

Association rule mining is a popular data mining method for finding relationships between objects or item sets [1] [2] [3]. Nowadays, association rule mining must include huge amounts of data. The Apriori method is popular for association rule mining. [4] We uncover frequent data patterns using the Apriori method. The Apriori algorithm explores (k+1) itemsets iteratively using k-item sets. First, scan the Database and count common 1-itemsets to locate frequent itemsets. Itemsets with minimum support are preserved. Use these to find frequent 2-itemsets. This continues until the freshly generated itemset is empty or until no further itemsets fulfil the minimal support criterion.

Association rules are determined by checking itemsets against a minimal confidence level. This technique must repeatedly scan the Database to generate frequent itemsets, which is difficult in the age of huge data. Figure 1 shows the Apriori algorithm flow chart. Apriori and other traditional association rule mining algorithms mine positive rules. Positive association rule mining detects items that are positively associated, meaning if one rises, the other rises also. The original use of positive association rule mining is market basket analysis, but it has been applied to biological datasets, web-log data, fraud detection, census data, and more. Items that are adversely associated fall if one goes up, according to negative association laws. Negative association rule mining can also be used to construct efficient crime data analysis decision support systems in healthcare [5].

Negative patterns are quite important, even more than positive patterns, due to the kind of impact that they create when such patterns exist. Negative patterns are often traced in the medical field, the financial field, and the forecasting sector. In the medical field, two drugs having different chemicals may contradict 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 is the focus of this research. The term negative indicates the absence of an item. It also indicates the contradiction between two or more item sets. In recent days, they are also identified by the name nonoverlapping patterns.

Negative association rules are generated when the correlation between two item sets is negative; the Confidence between the items is quite high, which is true even if the support between the Item sets is not higher than the threshold value. Some of the item sets could be regular but may need to appear together.

Patterns with a negative correlation among the Item sets imply that the occurrence of one item set has no relation with the occurrence of the other. When an Item set A or Item set B only occurs, and there are few transactions in which both A and B occur, the correlation between A and B is negative.

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

Association between the Frequent Patterns (Positive and Negative) Association rules are a precious outcome of pattern mining. Several algorithms have been developed to mine positive association rules. The term positive indicates the togetherness of the items wherever they appear in the transactional databases. The association rules are developed from positive patterns or item sets. If there exist 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 $A \Rightarrow B$.

Many associations exist, such as $(\neg A \Rightarrow \neg B)$, $(\neg A \Rightarrow B)$, $(A \Rightarrow \neg B)$, and $(A \Rightarrow B)$, which makes the problem of finding the negative associations much complicated. Many problems arise in finding the frequent item sets, infrequent item sets, proper positive and negative association rules, the problem caused by single minimum support, and so on.

Many mining methods exist in the literature for mining different types of patterns. The Mining methods that are in existence as of date can be classified into basic mining methods that include candidate generation methods that have been many (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 interestingness, including interestingness (subjective Vs objective), Constraint-based mining, mining correlation rules, and exception rules. The mining methods are also classified considering how the Database is organized, including a distributed, incremental, 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 data can be mined, including sequential, time series, Structural (Tree, lattice, graph), spatial (Colocation), temporal (evolutionary and periodic), Image, Video Multimedia and network patterns). Mining methods can be classified into pattern-based classification, clustering, semantic annotations, collaborative filtering, and Privacy-preserving.

The main problem in finding negative associations is handling huge databases and the right important negative associations. There is a need to consider closed and minimal item sets that are prominent, especially in the medical field so that focused investigations and corrections can be carried out.

Closed frequent Item sets substantially reduce the number of patterns generated in frequent Item set mining while preserving complete information regarding the set of frequent item sets. That is, the frequent item sets and the related support values can be easily derived from the set of closed item sets. It is more desirable to mine closed item sets rather than all set of all frequent item sets.

There are some approaches presented in the literature for mining negative associations. No method has been presented that aims at mining frequent, regular, Closed and Maximal item sets that are negatively associated and affected on medical databases, which is the focus of this research. This paper uses a Vertical format-based method for mining the Negative associations existing in the medical databases.

2. Problem Definition

In the medical field, negative associations are very dangerous to Human Beings. Wrong administration of the drugs will create many ill health situations than curing the real problem. It is necessary to find the contradicting drugs before the same is administered. The frequent and regular drugs, which are numerous, need special attention. There is a need to consider the closedness and Maximality of the itemset so that fewer frequent and regular item sets can be found and Negative associations are determined.

3. Objectives of the Research

1. To construct a Database and generate an example set which can be used to experiment to find the negative association among the regular, frequent, closed, and Maximal items.
2. To develop an algorithm that finds the Frequent, Regular, Closed, and Maximal item sets and finds negative associations among those item sets.
3. Find the most optimum threshold values for Frequency and Regularity, in which the most accurate negative associations can be found.

4. Representing Negative Associations

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

$$\text{Support}(i_1, \neg i_2, i_3) = \text{support}(i_1, i_3) - \text{sup}(i_1, i_2, i_3)$$

A rule of the form $(A \Rightarrow B)$ is a positive rule, and other forms of the rules $(\neg A \Rightarrow \neg B)$, $(\neg A \Rightarrow B)$, and $(A \Rightarrow \neg B)$ are negative rules.

The interestingness of a negative rule is the Confidence of the rule expressed in terms of

$$\text{sup}(A \cup \neg B) / \text{sup}(A)$$

The rules of the form $A \Rightarrow \neg B$ need to be discovered, which meets the minimum support and Confidence provided by the user with the condition that A and B are disjoint sets.

$$\text{—supp}(A) \geq \text{ms}, \text{supp}(B) \geq \text{ms} \text{ and } \text{supp}(A \cup B) < \text{ms};$$

$$\text{where ms} = \text{Minimum Support}$$

$$\text{—supp}(A \Rightarrow \neg B) = \text{supp}(A \cup \neg B);$$

$$\text{—conf}(A \Rightarrow \neg B) = \text{supp}(A \cup \neg B) / \text{supp}(A) \geq \text{mc}.$$

5. Related work

Ming-Syan Chen [6] has presented a comprehensive survey on the availability of different mining methods and the purpose for which the mining methods are used. A purposed-based classification of the mining methods and the purpose for which the mining methods are presented.

A basic method to mine transactional databases leads to mining too many patterns that will reflect into too many association rules that could be more interesting to the end user. Ashok Savasere [7] Presented a mining method that combines positive associations with domain knowledge so that very few negative associations are found, which can be easily evaluated and presented.

Balaji Padmanabhan [8] Has presented that pattern mining generally leads to too many patterns and needs to consider the domain knowledge that the decision-makers have. Decision makers have prior knowledge about the data regarding precepts and beliefs. They have also presented a method that mines unexpected patterns using beliefs and perceptions. They have experimented with the method using WEB log files and proved that efficient mining could be done using the user's perception. Many mining methods in the literature have used candidate generation approaches, which use an Apriori-like method. This approach has proved time-consuming and costly, especially when many long patterns are involved.

Jiawe Han [9] has proposed a novel method that uses a Frequent pattern tree (FP-tree), an extension of the structure. The FP-tree structure stores crucial information about the frequent patterns and the information used for pattern mining. The method used the concept of FP-growth for a complete mining set of frequent patterns.

The method proposed by Jiawe Han [9] used three main techniques, including database compression, a pattern fragment growth method and a divide and conquer method for decomposing mining tasks into small tasks that 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 horizontally or in vertical mining approaches. Vertical mining approaches are found to be quite effective when compared to horizontal approaches. Fast Frequency counting through intersecting operations on transaction IDs and pruning are not required, which are the main advantages of vertical format methods.

However, these methods need more memory when the entries to be made into a vertical format table need to be lighter. Mohammed J. Zaki [10] has presented a novel way of presenting the vertical data called Di-Set, which considers the differences between the transactions of candidate patterns and the very patterns themselves. They have shown how Disets can drastically cut down the memory requirement for storing 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 [11] that is used for generating both negative and positive association rules. The negative associations between the patterns were evaluated by checking the expressions like $(A \Rightarrow \neg B)$, $(\neg A \Rightarrow B)$, and $(\neg A \Rightarrow \neg B)$. The rules are mined from a large database by constraining the patterns using interesting patterns.

Some of the generated association rules could be exceptional because the rules are less interesting or have too high a confidence. Daly et al. [12] presented a method for evaluating exceptional mining rules. 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 to evaluate the patterns and decision-making.

Most of the methods proposed in the literature use interestingness measures to prune the most wanted patterns for decision-making. However, precisely defining the interestingness measure is 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 [13] 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 discovers the Most interesting rules presented.

Correlations are statistical measures that find how well a set of data records is related to another set of records. Maria-Luiza and Antonie [14] have presented a method to find negative association rules on the existence of a correlation between two item sets. Negative rules between the item sets are extracted if the correlation between the items is negative, and the Confidence of the item sets is quite high. The extracted negative association rules can have either consequents or antecedents ($\neg Y$ and Y) even when the computed support value from the item sets 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 a negative correlation exists between the Item sets, leading to negative association rules that negate both the antecedents and consequents. They have generated all negative and positive associations rules out of the patterns that strongly associate them. If no association rules are generated, the correlation threshold value will have to be lowered, reducing the strength of correlations between the items set.

A survey has been presented by Chris Cornelis [15] 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 catalogued several mining algorithms based on some parameters and could figure out the drawbacks of each. They have also presented a modified mining algorithm based on the Apriori approach that can find negative associations with interesting ones attached through a confidence framework.

They have used an upward closure property that conforms to the support-based interestingness of negative associations under validity definitions. Usually, the interestingness parameter "Support" is defined for entering the dataset. The data records are recognized as a hierarchy of records with a set of records at each level with a specific support value. Several support values are defined at each level.

A model proposed by [Xiangjun Dong [16] called MLMS (Multi-Level minimum support) considers defining minimum support value at each of the levels of the records. MLMS is used to discover both frequent and infrequent item sets. They have considered correlation and confidence interesting measures and proposed another interesting measure to mine both frequent and infrequent item sets. An algorithm called PNAR-MLMS has been proposed to generate positively and negatively associated patterns from frequent and infrequent item sets generated through the MNMS model.

Xiangjun Dong et al. [17] have also developed PNAR-based Classifiers using association rules classified into some known categories. The classifiers then can be used to find whether a pattern leads to a negative or positive association. Discovering the K-Most intersecting rule requires minimum support and threshold values. It is rather difficult to define the minimum threshold hold value as the user needs to know 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, is also presented in the literature. It requires the user to define the measure of interestingness and the number of rules the user is interested in. Xiangjun Dong [18] has extended the GRD method, which is used as a form for mining positive and negative rules. Both positive and negative association rules are mined through transactions. Negative Association rules explain how one pattern negates some other patterns.

Many applications exist that need the mining of the negative association rules, especially those that will help carry out the market-basket analysis. The negative association rules can also be used to develop classifiers, using which classification models are built. Mining negative association rules require the exploration of large data space. Many of the algorithms proposed in the literature need to be in use due to the usage of large databases.

Xiangjun Dong et al. [19] Have extended the support confidence framework by adding a correlation coefficient threshold that keeps sliding as data access moves. They have used correlation coefficients that can be calculated considering different patterns. Positive and negative correlated patterns are found in antecedents and consequents.

Most of the work is focused on frequent items till 2009, and only focus is made such time on regular items that occur in regular intervals. Regular items are the most important than frequent items, the occurrence of which has no time limitation. Tanbeer et al. [20] are the first authors to focus on regular item sets. They have proposed a "Regular pattern tree," a tree structure to discover regular patterns. The algorithm scans the Database twice. In the first set, the Regularity and support values of the item sets are determined, and in the second scan, a regular Pattern tree is constructed. The process adopted by them is like cyclic and periodic patterns. In many transactional databases, data is 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 the defined minimum support threshold the users define [21]. The minimum support threshold assumes that all frequent sequences have the same Frequency, which is not true 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 [22]. Has presented new models for mining interesting negative and positive associations among transactional data records. They have considered merging two algorithms, including interesting negative and positive association rules (PNAR) and mining interesting multiple-level support algorithms (IMAMS). Their proposed algorithm helps to mine positive and negative association rules from interesting frequent and in-frequent item sets using multiple support values.

NVS Pavan [23], [24] have presented a method of finding both positive and negative associations considering the Regularity of the Item set using the vertical Table mining method.

Yanqing Ji et al. [25] have focused on mining item sets that have casual relationships, which help either prevent or correct negative outcomes which are caused by the antecedents. They have presented an interesting new measure called exclusive casual leverage based on the RPD model (Computational, Fuzzy recognition prime decision model). The mining algorithm they used considered the Database that connects the drugs and the adverse reactions. They have, however, ignored the issue of Regularity and Maximality.

Bagui and Dhar et al. [26] have presented a method to mine positive and negative association rules, considering that various data is stored in a MAP REDUCE environment. They have used frequent item set mining using the Apriori algorithm, which has proved efficient due to creating many item sets and leading to heavy computing time requirements.

Few studies have examined negative association rule mining [27–31], but none in Biig data. Since positive association rule mining has been studied extensively,

Mahmood et al.[32] determined positive and negative association rules using infrequent item sets. Positive association rule mining extracts frequent things or item sets. However, it may discard many significant items or item sets with low support. Despite limited support, rare goods or item sets can elicit important negative association rules. Negative association rule mining is significant, but it requires more search space than positive association rule mining because low-support objects must be maintained. This would make sequential Apriori algorithm implementations easier and even harder on massive data. Negative association rule mining has been implemented a few times.

Brin et al. [33] and [34] suggested a Chi-square test for negative association rules. Positive and negative associations were determined using a correlation matrix. Mining strong collective item sets was Agrawal and Yu's method. They used positive frequent item sets and domain knowledge as a taxonomy to establish negative association rules. Taxonomy was utilized to pick negative item sets after all positive items were obtained. Selecting a negative itemset generated association rules. This domain-specific technique requires a predetermined taxonomy, making it hard to generalize. Similar methods are in [36].

Positive and negative association rules are found by Wu et al. [37]. The algorithm finds rules in the form of $\neg X \Rightarrow Y$, $X \Rightarrow \neg Y$, and $X \Rightarrow \neg Y$. With the support-confidence framework, authors included "mininterest". The dependency between two item sets was checked using main interest.

Substitution rule mining (SRM) by Teng et al. [38,39] finds a subset of negative association rules. The $X \Rightarrow \neg Y$ algorithm identifies negative association rules. This algorithm finds "concrete" elements first. Concrete things exceed projected support with a high Chi-square value. The correlation coefficient is calculated for each pair.

Antonie and Zaiane [40] developed a method that uses Pearson's ϕ correlation coefficient to identify strong positive and negative association rules. ϕ The correlation coefficient for $X \Rightarrow Y$

connection rule: GRD, Tiruvady, and Webb's [41] algorithm finds top-k positive and negative association rules. Use leverage and the amount of rules to discover. In [42], the authors proposed a new Apriori-based algorithm (PNAR) that finds negative association rules via upward closure. If $\neg X$ crosses the minimum support threshold, then $Y \subseteq I$, $X \cap Y = \emptyset$, and $\neg(XY)$ also meet it.

Md Saiful Islam et al. [43] have presented a comprehensive and systematic review of healthcare analytics using data mining and big data, following Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines. They have reviewed all the articles which were published during 2005 and 2016. The review conducted by them did not consider the issue of negative associations.

Hnin et al. [44] have used the Maximal frequent item set algorithm for mining item sets from a healthcare database. Relevant to heart diseases. A decision tree-based machine learning model is trained to learn and predict the occurrence of heart diseases. The data set required is mined by using a clustering algorithm. In this approach, the issue of Frequency, closed data sets, Regularity and negative associations have yet to be addressed.

SIMARJEET KAUR et al. [45] Have presented a review on using AI techniques for diagnosing the disease. No review, however, has been made on the drugs administered and the impact of the same when heart diseases are predicted.

JIANXIANG WEI et al. [46] have presented a Risk prediction model from drug reactions using machine learning approaches. They have predicted the risk of administering a drug for treating a disease. They have not considered any risk which arises when negatively associated drugs are administered to a Patient.

Lu Yuwen, Shuyu et al. [47] have proposed a framework that focuses on combining sequential data mining called "Prefix-Span" and a disproportionality-based method called "Proportional Report Ratio" to detect potential serious adverse drug reactions based on data collected related to casual relationships and drugs and drug reactions. They have considered single drug-to-drug reactions. They have not considered drug reactions to the use of constricting drugs, which have negative associations.

Yifeng Lu et al. [48] have presented that frequent item set mining less reveals expected patterns. The negative associations between the drugs recommended can thus be found. But the issue is infrequent item set, which also has negative associations among the drugs, is also crucial. The authors have presented a method for mining infrequent close item sets using bi-directional traversing. However, the negative associations among infrequent item sets or infrequent and frequent item sets have yet to be explored.

Jingzhuo Zhang et al. [49] have presented a method to extract interaction between the drugs administered to patients who are affected due to various kinds of diseases. They have developed a database of drug-drug interactions, considering various medical sources. They have applied distant supervision methods to extract drug-drug interactions. To extract the relationships between the drugs, a bidirectional encoder representation from transformers has been used. However, no modelling is done to classify whether the interactions are positive or negative.

Yanqing Ji et al. [50] have focused on mining causal relationships which happen rarely or the probability of occurrence of such item sets is rare. Discovering causal relationships and eliminating them will help prevent rare or negative outcomes. They have proposed a data mining framework in which a novel, interesting measure called exclusive casual leverage, based on a Computational fuzzy recognition-primed model, has been used to determine casual relationships. However, they still need to address the existence of Negative associations among regular and frequent item sets.

E. Ramaraj et al. [51] proposed an extended and modified Eclat method for finding positive and negative associations between frequent item sets. They have not considered either Regularity or rare item sets. They have not focussed on negative associations among the drugs based on chemical compositions or diseases based on drug reactions.

Wu et al. [52] presented an Apriori-based framework for mining positive and negative Association Rules. Another Apriori-based algorithm was given by Antonie and Zaki [53] to simultaneously

generate positive Association Rules and (a subclass of) negative Association Rules. Chris Cornelis et al. [54] proposed a new algorithm, SPNAR, for mining both positive and negative rules. They have also proposed the BAECLAT method for Mining Positive and Negative Rules from Large databases IEEE Conference 2006.

Mario Luiza Anionic et al. [55] have proposed a modified BAECLAT method for Mining Positive and Negative Association Rules through the process of confining the rules

Jigar R. Desai et al. [56] opined that there are underlying bios in the medical data, and no methods exist for handling uncertainty. They have proposed a method that accounts for the bios while identifying the associations between two rare genetic disorders and type 2 Diabetics (Diseases). They have considered both positive and negative control on the diseases and used negative control to estimate the extent of bios in several medical databases. The study did not focus on the negative associations among the drugs, considering their chemical compositions.

M. Goldacre et al. [57] have explained how large databases can be explored to identify the association between the diseases occurring commonly or less commonly than their frequencies. They have discussed some conditions associated with different diseases. They have shown an association between the conditions and reveal the association between the diseases. However, they have not discussed the type of association between the diseases.

Yoonbee Kim et al. [14,58] have proposed a method for constructing drug-gene-disease associations through generalized tensor decomposition. They used two networks created using chemical structures and ATC codes as drug features to predict the drug-gene-disease association. They learnt the features of the drugs, genes and diseases through learning a multi-layer perceptron-based neural network. They have considered all positive associations and not given much weight to negative associations, especially among the drugs.

The comparative analysis of various algorithms used, especially concerning Negative association mining, is shown in Table 1. It can be seen from the table the issue of maximality and closure have not been addressed while finding the negative association.

Table 1. Comparative Analysis of Mining Algorithms Concerning Negative Association Mining

[illegible]

6. Gap Analysis

Most existing studies are based on finding positive or negative associations considering the Frequency and the Regularity of the Item sets. When the Database is large, it leads to too many negative associations that do not matter much. It does not find the most critical negatively associated itemset. The choice of Maximality and closedness must be considered in addition to the Regularity and Frequency to arrive at the most significant Negative associations that matter much.

7. Methods to computing Negative Association among Frequent, regular, closed and Maximal Data Sets

Method-1

If Item sets X and Y are both frequent but rarely occur together, then $\text{sup}(X \cup Y) < \text{sup}(X) * \text{sup}(Y)$, and the pattern $X \cup Y$ is a negatively correlated pattern.

If $\text{sup}(X \cup Y) \ll \text{sup}(X) * \text{sup}(Y)$, then X and Y are strongly negatively correlated, and the pattern $X \cup Y$ is a strongly negatively correlated pattern. This definition can be extended to k item sets, But this has a Null transaction problem.

Method 2

Another way to compute negative associations is to calculate.

$\text{sup}(X \cup \bar{E}) * \text{sup}(\bar{A} \cup E) \gg \text{sup}(X \cup E) * \text{sup}(\bar{A} \cup \bar{E})$, which also has the problem of a null transaction problem

Method 3

Suppose that Item sets A and Y are frequent, i.e. $\text{sup}(A) \geq \text{min-sup}$, $\text{sup}(B) \geq \text{min-sup}$, where min-sup is the minimum support threshold.

Then $P(A \cup B) + P(B \cup A) / 2 < \epsilon$, where ϵ is the negative pattern threshold. This way of computing the negative association is free from the problem of null transactions.

8. Computing the Negative Associations from Regular, frequent, closed and Maximal Item Sets. -Algorithm

Algorithm-A

1. Read the support value that dictates the threshold value of the Frequency of the patterns and the Regularity defined by the user.

2. Read the data in the 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 Nonfrequent items.

5. Find the Closed Item sets and maximal Item sets and include them in Table 3

For every record in Table 2

{

Selecting Closed and Maximal Item Sets

Suppose the Item set is a subset of the existing data set in Table 2. Loop.

If the Item set in the record is a superset of any other record in Table 2 For every record in Table 3

If the Item set is a superset of a record in Table 3 with the same support, Prune the record in Table 3.

Add the record to Table 3 as a Close Maximal Item Set.

}

6. For every record in Table 3 For every Next Record in Table 3

i. Find the intersection of the current record and the next record.

ii. If the intersection is null, enter the current and next items into a negative table in Table 4.

iii. If the intersection is not null,

iv. Get the common elements.

- a. If the count of elements is > the regularity threshold and the frequency threshold, Add the Records to Table 3 at the end.
- b. LOOP if the common elements do not satisfy the regularity or frequency constraint.
7. Report the Negatively Associated Item Sets that meet the Frequency, Regularity, Maximality and Closedness.

9. Data Set for Experimentation

A database contains Patent Registration Details, Diagnosis Codes, Patient-Diagnosis Details, Chemical Codes, Drug-Chemical Details, Quantity Codes and Prescription details. 100,000 Patient Registration and the associated prescriptions have been collected from different hospitals and stored in the Database. An Example set has been generated containing the data related to each diagnosis, drugs administered, and the related Chemical composition of those drugs. Each data item in the repeated groups is encoded, and the data items are replaced with codes. The Example set is sorted, the Frequency and Regularity of the Item set are computed, the Database is updated, and 100,000 records have been imported in a Flat file structure. These records have been processed using the algorithm Proposed in this paper. No standard data set is available anywhere containing the data elements required for finding the existence of Negative Associations.

10. Results and Discussion

10.1. Results – Implementation of the Algorithm -A on the dataset

Step1: Extract Sample data from the Database

Algorithm A is implemented on a sample example set containing 10 Patients, 23 Diseases, 44 Drugs and 15 Chemical compositions. The details of the sample data selected are shown in Table 2.

Table 2. Sample Medical Data Extracted from the Database

PSL.No	Transaction ID	Patient Number	Disease	Drug	Chemicals					Drug	Chemicals			
1	T1	P100	DE1	DR1	CH1	CH2	CH3	NA	NA	DR2	CH4	CH5	CH9	CH10
	T2	P100	DE2	DR3	CH4	CH5	CH6	NA	NA	DR4	CH10	CH15	NA	NA
	T3	P100	DE3	DR5	CH2	CH3	CH7	NA	NA	DR6	CH13	CH14	CH15	NA
2	T4	P223	DE4	DR7	CH5	CH8	CH10	NA	NA	DR8	CH11	CH15	NA	NA
	T5	P223	DE5	DR9	CH1	CH3	CH5	CH16	CH19	NA	NA	NA	NA	NA
3	T6	P749	DE6	DR10	CH4	CH5	CH16	CH19	NA	NA	NA	NA	NA	NA
4	T7	P937	DE7	DR11	CH2	CH3	CH7	CH11	NA	DR12	CH12	CH13	NA	NA
5	T8	P119	DE8	DR13	CH5	CH8	CH11	NA	NA	DR14	CH12	CH14	CH15	NA
	T9	P119	DE9	DR15	CH1	CH3	CH5	NA	NA	DR16	CH8	CH9	NA	NA
	T10	P119	DE10	DR17	CH2	CH3	CH7	CH8	NA	DR18	CH13	CH14	CH15	NA
6	T11	P1235	DE11	DR19	CH5	CH8	CH11	CH15	NA	DR20	NA	NA	NA	NA
7	T12	P11	DE12	DR21	CH4	CH5	CH6	NA	NA	DR22	CH10	CH15	NA	NA
	T13	P11	DE13	DR23	CH2	CH3	CH7	CH8	NA	DR24	CH13	CH14	CH15	NA
	T14	P11	DE14	DR25	CH5	CH8	CH11	CH15	NA	DR26	NA	NA	NA	NA
8	T15	P4573	DE15	DR27	CH1	CH3	CH5	NA	NA	DR28	CH9	CH11	NA	NA
	T16	P4573	DE16	DR29	CH4	CH5	CH6	NA	NA	DR30	CH14	CH15	NA	NA
9	T17	P8765	DE17	DR31	CH2	CH3	CH6	CH7	NA	DR32	CH12	CH13	NA	NA
	T18	P8765	DE18	DR33	CH5	CH8	CH11	CH12	NA	DR34	CH14	CH15	NA	NA
10	T19	P10987	DE19	DR35	CH1	CH3	CH5	NA	NA	DR36	CH6	CH9	CH10	NA
	T20	P10987	DE20	DR37	CH4	CH5	CH6	NA	NA	DR38	CH12	CH14	CH15	NA
	T21	P10987	DE21	DR39	CH2	CH3	CH4	NA	NA	DR40	CH7	CH13	NA	NA
	T22	P10987	DE22	DR41	CH5	CH8	CH11	NA	NA	DR42	CH12	CH15	NA	NA
	T23	P10987	DE23	DR43	CH1	CH3	CH5	NA	NA	DR44	CH9	CH14	NA	NA

P - Patient, DE - Disease, DR = Drug, CH - Chemical in the Drug
Transaction IDS added to the data set to start-width. The first 20 records are shown in Table 2.
Here, Frequency is the count of transactions in which the item set appears.

Step-2 Add Transaction IDs

Transaction IDs are Added to Table 1 to keep track of each entry in the Table.

Step 3 Convert Table 1 into a vertical format

Table 1 is converted into a Vertical format showing the Occurrence of Each Chemical in Different Transactions. Only chemicals are considered related to the drugs used on the Patients. The Regularity and Frequency of the Items is computed as shown in **Table 2**

Step 4 Prune the Records which do not meet the threshold levels of Regularity and Frequency.

The Maximum Regularity (4) and the Minimum. Frequency (3), recommended by the Users, is used to Prune the Records that do not meet the threshold defined by the Users. **Table 3** shows that the Chemical codes CH2, CH10, CH12, CH14 and CH15 have been pruned as they do not meet the threshold value relating to Regularity and Frequency. The records left over are shown in Table. 4. Using this criterion, 5 Chemicals have been pruned.

Step 4 Prune the records which satisfy the Maximality and the Closedness criteria.

CH1 is a subset of CH3, CH7 is a subset of CH3, CH9 is a subset of CH5, and CH13 is a subset of CH3. Therefore, the Records are pruned. The leftover records after pruning are shown in **Table 4**.

Step 5 Find the Negatively Associated Chemicals

Find the Records with no common Transactions (Nill Common Items) that form the Negative Associations. Application of Intersection on the records Yields negative associations such as

(CH4 \Rightarrow CH8), (CH4 \Rightarrow CH11), (CH4 \Rightarrow CH8, CH11), (CH6 \Rightarrow CH8), (CH6 \Rightarrow CH11), (CH6 \Rightarrow CH8, CH11), (CH8 \Rightarrow CH4, CH6), (CH11 \Rightarrow CH4, CH6), (CH4, CH6), \Rightarrow CH8, CH11)

Table 3. Inverted Table

Chemical Code	Transaction Ids												Maximum Regularity (4)	Minimum Frequency (3)				
CH1	T1	T5	T9	T13	T17	T21							4	6				
CH2	T1	T3	T7	T11	T5	T9							6	6				
CH3	T1	T3	T5	T7	T9	T11	T13	T15	T17	T19	T21				2	11		
CH4	T1	T2	T6	T10	T14	T18	T19							4	7			
CH5	T1	T2	T4	T5	T6	T8	T9	T10	T12	T13	T14	T16	T17	T18	T20	T21	2	16
CH6	T2	T5	T6	T10	T14	T15	T17	T18							4	8		
CH7	T3	T7	T11	T15	T19							4	5					
CH8	T4	T8	T9	T11	T12	T16	T20							4	7			
CH9	T1	T5	T9	T13	T17	T21							4	6				
CH10	T1	T2	T4	T10	T17							7	5					
CH11	T4	T7	T8	T12	T13	T16	T20							4	7			
CH12	T7	T8	T15	T16	T18	T20							7	6				
CH13	T3	T7	T11	T15	T19							4	5					
CH14	T1	T3	T8	T11	T14	T16	T18	T21				5	8					
CH15	T2	T3	T4	T6	T8	T10	T12	T14	T16	T18	T20				7	11		

Table 4. List of Over-item sets after pruning based on Maximum Regularity and Minimum Frequency

Chemical Code	Transaction Ids																Maximum Regularity (4)	Minimum Frequency (3)		
CH1	T1	T5	T9	T13	T17	T21											4	6		
CH3	T1	T3	T5	T7	T9	T11	T13	T15	T17	T19	T21						2	11		
CH4	T1	T2	T6	T10	T14	T18	T19											4	7	
CH5	T1	T2	T4	T5	T6	T8	T9	T10	T12	T13	T14	T16	T17	T18	T20	T21	2	16		
CH6	T2	T5	T6	T10	T14	T15	T17	T18											4	8
CH7	T3	T7	T11	T15	T19													4	5	
CH8	T4	T8	T9	T11	T12	T16	T20												4	7
CH9	T1	T5	T9	T13	T17	T21												4	6	
CH11	T4	T7	T8	T12	T13	T16	T20												4	7
CH13	T3	T7	T11	T15	T19													4	5	

Table 5. List of item sets left over after pruning based on the Closedness and Maximality

Chemical Code	Transaction Ids																Maximum Regularity (4)	Minimum Frequency (3)		
CH3	T1	T3	T5	T7	T9	T11	T13	T15	T17	T19	T21						2	11		
CH4	T1	T2	T6	T10	T14	T18	T19											4	7	
CH5	T1	T2	T4	T5	T6	T8	T9	T10	T12	T13	T14	T16	T17	T18	T20	T21	2	16		
CH6	T2	T5	T6	T10	T14	T15	T17	T18										4	8	
CH8	T4	T8	T9	T11	T12	T16	T20												4	7
CH11	T4	T7	T8	T12	T13	T16	T20												4	7

Step 6 Find Negatively Associated Drugs

Map back to the Chemicals associated with the Drugs and find the negatively associated drugs, as shown in Table 6

Table 6. Mapping Negatively Associated Chemicals to the Drugs

Chemical	Associated Drug	Chemical	Associated Drug	Chemical	Associated Drug	Chemical	Associated Drug
CH4	DR3	CH8	DR7	CH11	DR11		
CH6	DR3	CH8	DR7	CH11	DR11		
CH4	DR3	CH6	DR3	CH8	DR7	CH11	DR11

adjust width These negative associations reveal that DR3 should not be used with DR7 or DR11 as both contradict each other.

11. Discussion

The 100,000 Examples have been Created through data collection and analyzed for different sample sizes drawn of sizes 30,000. 50,000 and 70,000 from the perspectives of different Regularity and Support, and the Number of Negative Associations found through applying support, regularity, Maximality and Closedness criteria.

Table 7 shows, considering 30,000 examples, the number of frequent, regular negative associations and the number of frequent, regular, closed, and Maximal negative associations which have been mined using Algorithm A. The negative associations have been generated by Keeping the Regularity fixed and varying the support.

Table 7. Analysis of Negative Frequent Regular Itemset with 30,000 Examples

Total Transactions	%Max Regularity	%Support Count	Number of Negative Frequent Regular	Number of Negative Frequent Regular Maximal and closed Items
30,000	3	2	6	2
	3	1.75	42	13
	3	1.625	154	46
	3	1.5	461	138
30,000	2.5	1.75	41	12
	2.5	1.625	154	46
	2.5	1.5	352	106
	2.5	1.125	981	294
30,000	2	1.75	35	11
	2	1.625	118	35
	2	1.5	181	54
	2	1.125	334	100

Table 8 shows the variation in the Number of Negative Associations, fixing the Frequency @ 1.75 and varying the Regularity between 2 and 3 considering 30,000 examples. Figure 1 shows the line graphs separately for the criteria (Frequency, Regularity) and (Frequency, Regularity, Closed and Maximality). On average, the number of negative associations gets reduced by 46.90% when the closedness and Maximal criteria are also considered.

Table 8. Number of Negative Associations considering two criteria, fixing the Frequency at 1.75 and Varying Regularity when 30,000 examples are selected.

At Frequency 1.75 and Recs = 30,000			
Regularity	Number of Negative Frequent Regular Item Sets	Number of Negative Frequent Regular Item, Closed and Maximal Item sets	% Decrease
3.00	6	2	66.7
2.50	41	27	34.1
2.00	35	21	40.00
Average % of decrease of Negative Associations			46.90

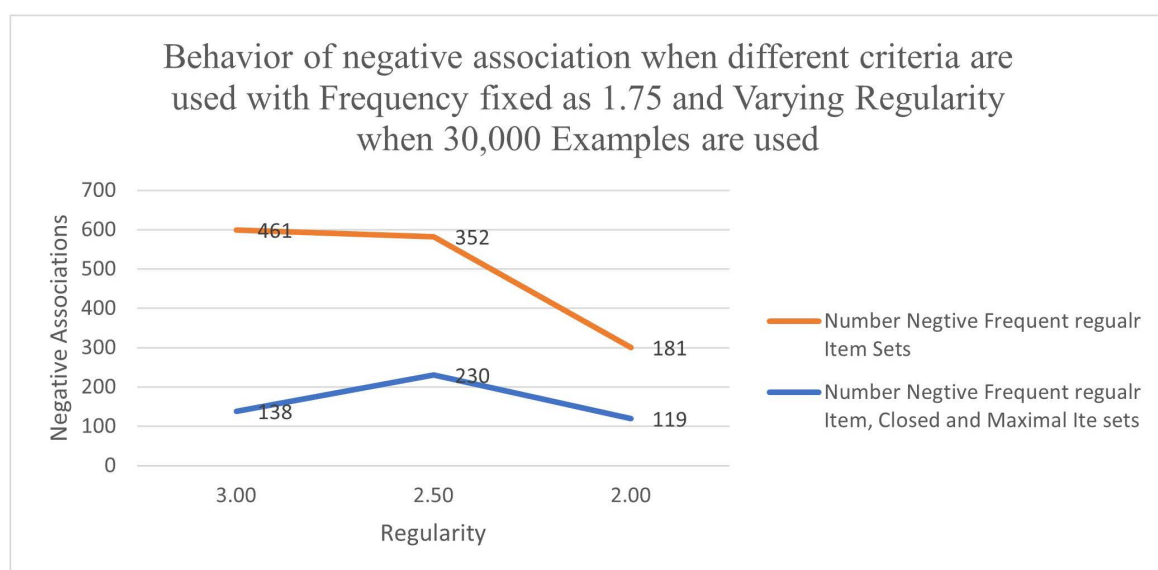
**Figure 1.** Variance of Negative Association Item sets fixing the Frequency (1.75) and varying the Regularity for 30,000 Example Records.

Table 9 shows the variation in the Number of Negative Associations, fixing the Frequency @ 1.5 and varying the Regularity between 2 and 3 considering 30,000 examples. Figure 2 shows the line graphs separately for the criteria (Frequency, Regularity) and (Frequency, Regularity, Closed and Maximality). On average, the number of negative associations gets reduced by 46.30% when the closedness and Maximal criteria are also considered.

Table 9. Number of Negative Associations considering two criteria, fixing the Frequency at 1.5 and Varying Regularity when 30,000 examples are selected.

At Frequency 1.5 and Recs = 30,000			
Regularity	Number of Negative Frequent Regular Item Sets	Number of Negative Frequent Regular Item, Closed and Maximal Item sets	% Decrease
3.00	461	138	0.701
2.50	352	230	0.347
2.00	181	119	0.343
Average			0.463

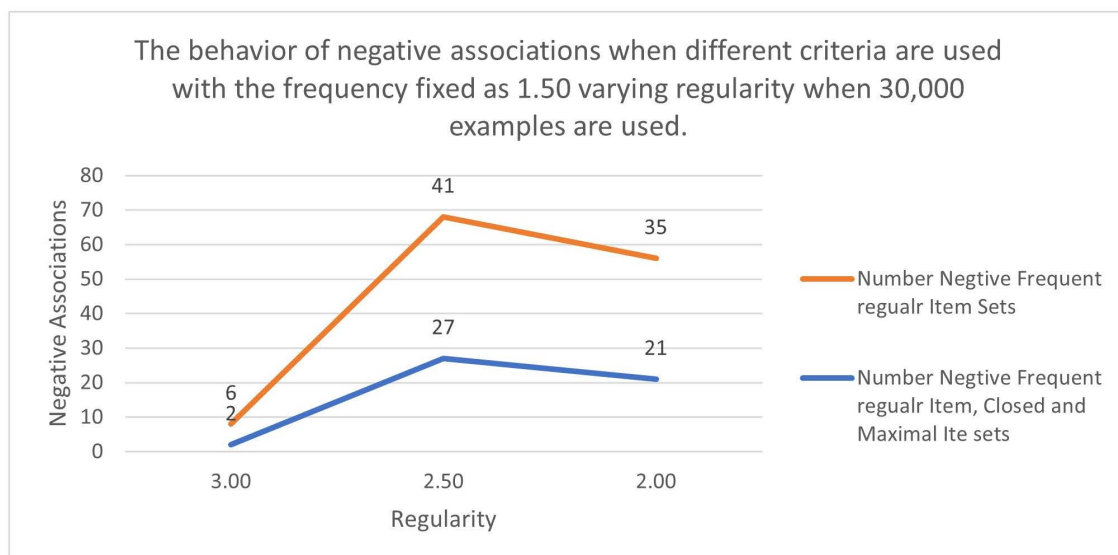


Figure 2. Number of Negative Associations considering two criteria, fixing the Frequency at 1.5 and Varying Regularity when 30,000 examples are selected.

Further Analysis has been carried out considering the higher size of the example set. Table 9 shows the reduction percentage of negative associations as the number of Examples used increases (30,000, 50,000, 70,000, 80,000). The % reduction in Negative association could be fixed at 70% by choosing the appropriate Frequency and Regularity for a sample size.

Table 10. % reduction in Negative association with increase in sample sizes and selection of suitable Frequency and support

Total Transactions	%Max Regularity	%Support Count	Number of Negative Frequent Regular	Number of Negative Frequent Regular Maximal and closed Items	Reduction in Negative associations	% Reduction
30,000	1.50	1.750	2	0	2	100
	1.50	1.625	3	1	2	67
	1.50	1.500	3	1	2	67
50,000	1.65	1.65	13	3	10	77
	1.65	1.25	41	10	31	76
	1.65	1.00	150	50	100	67
70,000	1.35	1.65	35	15	20	57
	1.35	1.35	118	35	83	70
	1.35	1.00	181	54	127	70
80,000	1.00	0.875	35	15	20	57
	1.00	0.815	118	35	83	70
	1.00	0.75	181	54	127	70
Average Improvement						71

Further Analysis is carried out to study the effect of fixing the Regularity and Varying Frequency and considering different sample sizes. Table 11 shows the number of negative associations fixing Regularity fixed at 1.50 and the sample size fixed at 30,000. The negative associations were reduced by about 73%. Figure 3 shows the variations.

Table 11. Analysis of variations in Negative associations as the Frequency

Frequency	Number Negative Frequent Regular Item Sets	Number Negative Frequent Regular Item, Closed and Maximal Item sets	% reduction
1.750	2	0	1.00
1.625	3	1	0.67
1.500	3	1	0.67
Average			0.73

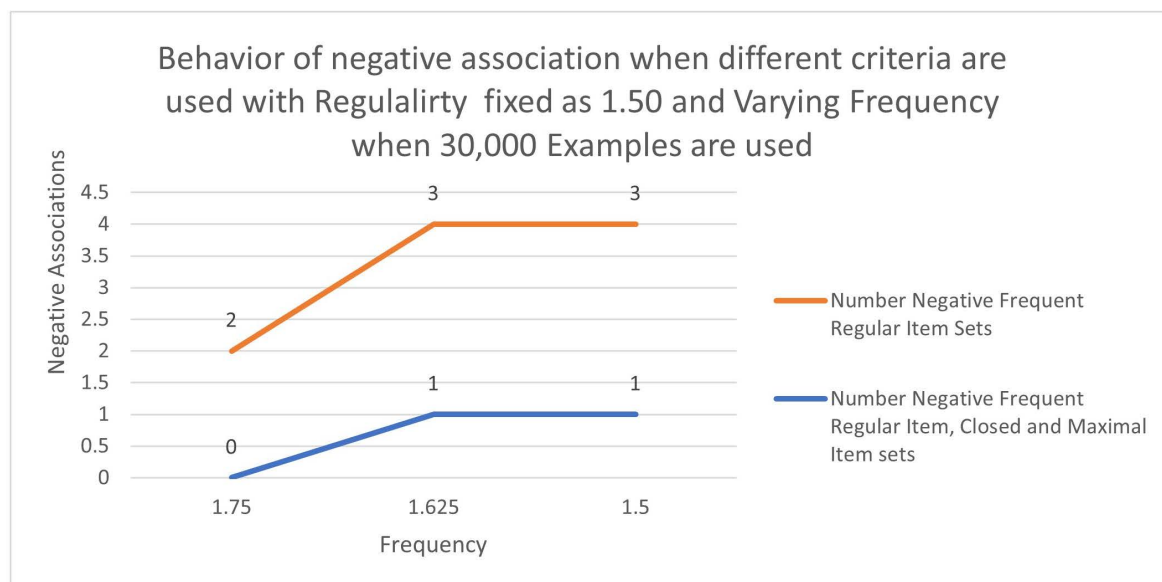
**Figure 3.** Variance of Negative Association Item sets fixing the Regularity (1.50) and varying the Frequency for 30,000 Example Records.

Table 12 shows the number of negative associations fixing Regularity fixed at 1.65 and the sample size fixed at 50,000. The negative associations were reduced by about 73%. Figure 4 shows the variations.

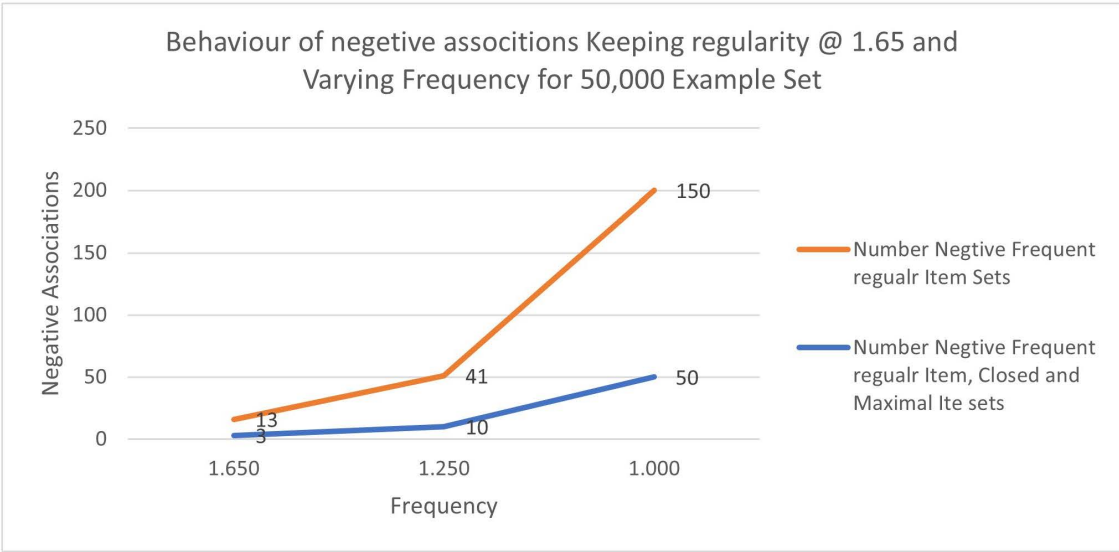


Figure 4. Variance of Negative Association Item sets fixing the Regularity (1.65) and varying the Frequency for 50,000 Example Records.

Table 12. Number of Negative associations fixing Regularity @ 1.65, No of Records @ 50,000 and varying Frequency.

Frequency	Number Negative Frequent Regular Item Sets	Number Negative Frequent Regular Item, Closed and Maximal Item sets	% reduction
1.650	13	3	0.77
1.250	41	10	0.76
1.000	150	50	0.67
Average			0.73

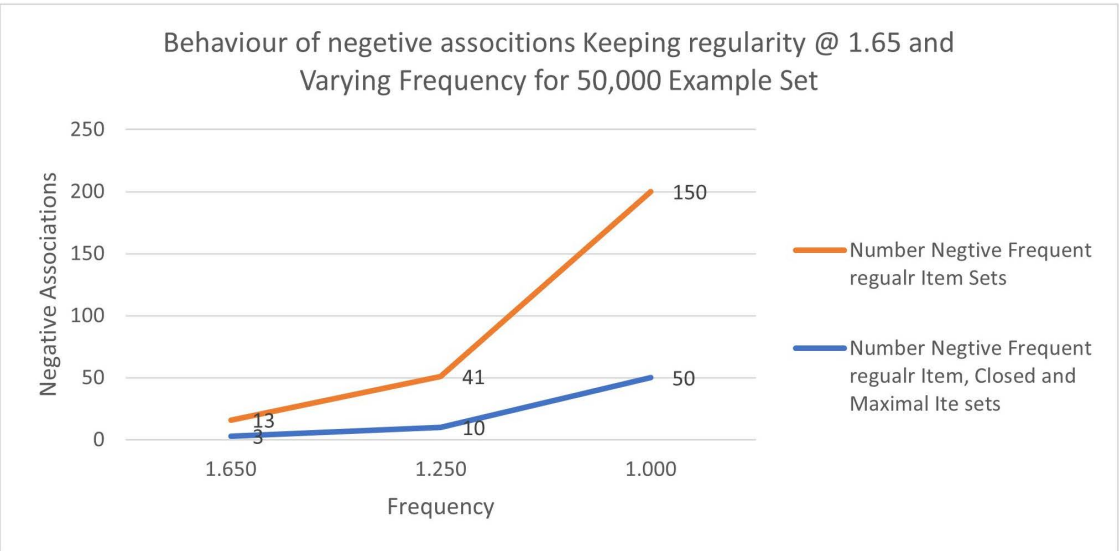


Figure 5. Variance of Negative Association Item sets fixing the Regularity (1.65) and varying the Frequency for 50,000 Example Records.

We can see that in both cases (30,000 Examples and 50,000 Examples), the percentage reduction in negative associations is about 73%. It can be concluded by fixing the Regularity and varying

Frequency. Yields 73% compared to a 43% reduction when Frequency is fixed, and the Regularity is varied. Variance in Frequency is the key issue when finding negative associations.

12. Conclusions

Finding negative associations among the drugs administered to patients due to the diseases that affected them.

Many have suffered during the Corona period due to too many side reactions caused due to administering drugs to cure Corona. Remedifier drugs, when given with other drugs, create several side reactions such as Black Fungus, heart attacks and many more. When drugs are chosen, finding whether they form negative associations is important. This research helps find all the negative associations among a set of drugs decided to be administered to the Patients. The number of negative associations could be very high, making it difficult to find the right ones.

The negative associations could be reduced to about 73% using the algorithm presented in this paper, considering fixing the Regularity and varying Frequency and applying the principality of Maximality and closedness. The remaining 27% are the crucial ones, which give the miss configurations of the drugs.

Every drug has a chemical composition. Intermixing of the chemical compositions sometimes gives negative results. Considering negative associations among the chemicals and converting those negative associations decoded into drugs will help find the most crucial and critical negative associations.

13. Future Scope

Further research can be carried out considering the constraints and colossal patterns that should be imposed when administering certain types of drugs.

References

1. Aggarwal CC, Yu PS. Mining associations with the collective strength approach. *IEEE Trans Knowl Data Eng.* 2001;13(6):863–73.
2. Aggarwal CC, Yu PS. A new framework for item-set generation. In: *Proceedings of the seventeenth ACM SIGACTSIGMOD-SIGART symposium on principles of database systems, PODS'98.* 1998. p. 18–24.
3. Agrawal R, Imielinski T, Swami A. Mining association rules between sets of items in large databases. In: *ACM SIGMOD conference.* New York City: ACM Press; 1993. p. 207–16.
4. Agrawal R, Srikant R. Fast algorithms for mining association rules. In: *VLDB 1994 proceedings of the 20th international conference on large databases.* 1994. p. 487–99.
5. Mahmood S, Shahbaz M, Guergachi A. Negative and positive association rules mining from text using frequent and infrequent itemsets. *Sci World J.* 2014;2014:1–11.
6. 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>
7. 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.
8. 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.
9. Jiawei 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>
10. Mohammed J. Zaki, *Fast Vertical Mining Using Diffsets*, SIGKDD 2003, ACM Copyright 2003.
11. 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>.

12. 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.
13. 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
14. 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
15. 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>.
16. 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 Zaiane, O.R., Li, Z. (eds.) ADMA 2006. LNCS (LNAI), vol. 4093, pp. 100–109. Springer, Heidelberg (2006), https://doi.org/10.1007/11811305_10
17. 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>
18. 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
19. Xiangjun Dong, Zhendong Niu, Donghua Zhu, Zhiyun Zheng4, 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
20. TANVEER 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
21. 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>
22. 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)
23. 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
24. 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.
25. Yanqing Ji, Hao Ying, Fellow, IEEE, John Tran, Peter Dews, Ayman Mansour, and R. Michael Massanari, A Method for Mining Infrequent Causal Associations and Its Application in Finding Adverse Drug Reaction Signal Pairs, IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 25, NO. 4, APRIL 2013
26. Bagui and Dhar, Positive and negative association rule mining in Hadoop's MapReduce environment, J Big Data (2019) 6:75 <https://doi.org/10.1186/s40537-019-0238-8>
27. Jiang H, Luan X, Dong X. Mining weighted negative association rules from infrequent itemsets based on multiple support. In: 2012 international conference on industrial control and electronics engineering. 2012. p. 89–92.
28. Kishor P, Porika S. An efficient approach for mining positive and negative association rules from large transactional databases. In: 2016 international conference on inventive computation technologies (ICICT). 2016.
29. Ramasubbareddy B, Govardhan A, Ramamohanreddy A. Mining positive and negative association rules. In: 5th international conference on computer science and education, Hefei, China, 2018. p. 1403–6.
30. Sahu AK, Kumar R, Rahim N. Mining negative association rules in a distributed environment. In: 2015 International conference on computational intelligence and communication networks. 2015. p. 934–7.

31. interesting frequent and infrequent itemsets. In: 2012 9th international conference on fuzzy systems and knowledge discovery. 2012. p. 650–5.
32. Mahmood S, Shahbaz M, Guergachi A. Negative and positive association rules mining from text using frequent and infrequent itemsets. *Sci World J.* 2014;2014:1–11
33. Brin S, Motwani R, Silverstein C. Beyond market basket: generalizing association rules to correlations. In: *Proc. SIGMOD.* 1997. p. 265–76.
34. Antonie L, Li J, Zaiane OR. Negative association rules. In: *Frequent pattern mining.* Berlin: Springer; 2014. p. 135–45. Bagui and Dhar J *Big Data* (2019)
35. Savasere A, Omiecinski E, Navathe S. Mining for strong negative associations in a large database of customer transactions. In: *Proc. of ICDE.* 1998. p. 494–502.
36. Yuan X, Buckles BP, Yuan Z, Zhang J. Mining negative association rules. In: *Proc. of ISCC.* 2002. p. 623–8
37. Wu X, Zhang C, Zhang S. Efficient mining of positive and negative association rules. *ACM Trans Inf Syst.* 2004;22(3):381–405
38. Teng W-G, Hsieh M-J, Chen M-S. On the mining of substitution rules for statistically dependent items. In: *Proc. of ICDM.* 2002. p. 442–9.
39. Teng W-G, Hsieh M-J, Chen M-S. A statistical framework for mining substitution rules. *Knowl Inf Syst.* 2005;7(2):158–78.
40. Antonie M-L, Zaiane OR. Mining positive and negative association rules: an approach for confined rules. In: *Proc. of PAKDD.* 2004. p. 27–38.
41. Thiruvady DR, Webb GI. Mining negative rules using GRD. In: *Proc. of PAKDD.* 2004. p. 161–5.
42. Cornelis C, Yan P, Zhang X, Chen G. Mining positive and negative association rules from large databases. In: *Proc. of CIS.* 2006. p. 1–6.
43. Md Saiful Islam 1, Md Mahmudul Hasan 1, Xiaoyi Wang 1, Hayley D. Germack 1,2,3 and, Md Noor-E-Alam 1,*A Systematic Review on Healthcare Analytics: Application and Theoretical Perspective of Data Mining, *Healthcare*, 2018, 6, 54; doi:10.3390/healthcare6020054
44. Hnin Wint Khaing, Data Mining based Fragmentation and Prediction of Medical Data, 978-1-61284-840-2/11/2011 IEEE
45. SIMARJEET KAURI, JIMMY SINGLA 1, LEWIS NKENYEREYE 2, SUDAN JHA 3, (Senior Member, IEEE), DEEPAK PRASHAR1, (Member, IEEE), GYANENDRA PRASAD JOSHI4, SHAKER EL-SAPPAGH5,6, MD. SAIFUL ISLAM 7, AND S. M. RIAZUL ISLAM, Medical Diagnostic Systems Using Artificial Intelligence (AI) Algorithms: Principles and Perspectives, *IEEE Access* Volume 8, 2020
46. JIANXIANG WEI 1, ZHIQIANG LU 2, KAI QIU 2, PENGYANG LI 2, AND HAOFEI SUN 2, Predicting Drug Risk Level from Adverse Drug Reactions Using SMOTE and Machine, Learning Approaches, *IEEE Access*, VOLUME 8, 2020, 185761- 185775
47. Lu Yuwen, Shuyu Chen and Hancui Zhang, Detecting Potential Serious Adverse Drug Reactions using Sequential Pattern Mining Method, 978-1-5386-6565 -7118
48. Yifeng Lu, Thomas Seidl, Towards Efficient, Closed Infrequent Itemset Mining using Bi-directional Traversing, 2018 IEEE 5th International Conference on Data Science and Advanced Analytics,
49. Jingzhuo Zhang, Weijie Liul and Ping Wangl,2,3,*, Drug-Drug Interaction Extraction from Chinese Biomedical Literature, using distant supervision, 2020 IEEE International Conference on Knowledge Graph (ICKG), 978-1-7281-8156-1/20/ ©2020 IEEE, DOI 10.1109/ICBK50248.2020.00089, 593-598
50. Yanqing Ji, Hao Ying, Fellow, IEEE, John Tran, Peter Dews, Ayman Mansour, and R. Michael Massanari, A Method for Mining Infrequent Causal Associations and Its Application in Finding Adverse Drug Reaction Signal Pairs, *IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING*, VOL. 25, NO. 4, APRIL 2013, 721-733
51. E.Ramaraj, N Venkatesan, Positive and Negative Association Rule Analysis in Health Care Database, *IJCSNS International Journal of Computer Science and Network Security*, VOL.8 No.10, pp. October 2008, 325-330
52. X. Wu, C. Zhang and S. Zhang, "Efficient Mining of Both Positive and Negative Association Rules", *ACM Trans. on Information Systems*, vol. 22(3), 2004, pp 381–405.
53. Chris Cornelis, Peng Yan, Xing Zhang, Guoqing Chen Mining Positive and Negative Rules from Large databases IEEE conference 2006.
54. M.L. Antonie and OR Zaiane, "Mining Positive and Negative Association Rules: an Approach for Confined Rules", *Proc. Intl. Conf. on Principles and Practice of Knowledge Discovery in Databases*, 2004, pp 27–38.

55. Mario Luiza Antonic, Osmar R. Zaiane Mining Positive and Negative Association Rules – an approach for confine rules
56. Jigar R. Desai, Craig L. Hyde, Shaum Kabadi, Matthew St Louis, Vinicius Bonato, A. Katrina Loomis, Aaron Galaznik and Marc L. Berger, Utilization of Positive and Negative Controls to Examine Comorbid Associations in Observational Database Studies, *Medical Care*, Vol. 55, No. 3 (March 2017), pp. 244-251 (8 pages), Published By: Lippincott Williams & Wilkins
57. M. Goldacre, L. Kurina, D. Yeates, V. Seagroatt, L. Gill, Use of large medical databases to study associations between diseases, *QJM: An International Journal of Medicine*, Volume 93, Issue 10, October 2000, Pages 669–675, <https://doi.org/10.1093/qjmed/93.10.669>
58. Yoonbee Kim 1 and Young-Rae Cho 1,2, Predicting Drug–Gene–Disease Associations by Tensor Decomposition for Network-Based Computational
59. Drug Repositioning, *Biomedicines* 2023, 11, 1998. <https://doi.org/10.3390/biomedicines11071998>

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.