

Article

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

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

¹ Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Guntur 522302, Andhra Pradesh, India; rajaraob@yahoo.com

² Department of Electronics and Computer Science, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Guntur 522302, Andhra Pradesh, India

* Correspondence: drsastry@kluniversity.in

Abstract: Many data mining studies have focused on mining positive associations among frequent and regular item sets. However, none have considered time and regularity bearing in mind such associations. The frequent and regular item sets will be huge, even when regularity and frequency are considered without any time consideration. Negative associations are equally important in medical databases, reflecting considerable discrepancies in medications used to treat various disorders. It is important to find the most effective negative associations. The mined associations should be as small as possible so that the most important disconnections can be found. This paper proposes a mining method that mines medical databases to find regular, frequent, closed, and maximal item sets that reflect minimal negative associations. The proposed algorithm reduces the negative associations by 70% when the maximal and closed properties have been used, considering any sample size, regularity, or frequency threshold.

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



Citation: Budaraju, R.R.; Jammalamadaka, S.K.R. Mining Negative Associations from Medical Databases Considering Frequent, Regular, Closed and Maximal Patterns. *Computers* **2024**, *13*, 18. <https://doi.org/10.3390/computers13010018>

Academic Editor: Paolo Bellavista

Received: 13 November 2023

Revised: 29 December 2023

Accepted: 3 January 2024

Published: 8 January 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Association rule mining is a popular data mining method for finding relationships between objects or item sets [1–3]. 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 $(k + 1)$ item sets iteratively using k -item sets. First, scan the database and count common one-item sets to locate frequent item sets. Item sets with minimal support are preserved. Find common two-item sets with these. This continues until the newly created item set is empty or no longer meets the minimal support condition. Association rules are determined by checking item sets against a minimal confidence level. This technique must repeatedly scan the database to generate frequent item sets, which is a difficult task in this age of big data. Apriori and other traditional association rule mining algorithms mine positive rules. Positive association rule mining detects positively associated items, meaning that if one rises, the other also rises. The positive association rule mining has been applied to web log data, biological data sets, census data, fraud detection, and more. According to negative association laws, adversely associated items fall if one goes up. Negative association rule mining can also be used to construct efficient crime data analysis decision support systems in healthcare, etc. [5].

Negative patterns are more essential than positive ones because of their influence. Negative patterns are common in finance, medicine, and prediction. Two medications with distinct ingredients may conflict in medicine. A temperature rule may not apply to a cool zone. Discovering frequent, regular, and unfavourable trends is most significant in this

research. Negative means absent. It also implies conflict between two or more item sets. Recently, these have been called non-overlapping patterns.

Negative association rules form when two item sets have a negative correlation and high confidence, even if the support is below the threshold value. Some regular item sets need to appear together. Patterns with a negative correlation among the item sets imply that the occurrence of one item set has no relationship with the occurrence of the other. A and B are negatively correlated when only one of A or B occurs, and few transactions occur with both. Unique patterns might involve negative terms and be considered as negative rules. These patterns are association rule exceptions called unexpected patterns.

The relationship between positive and negative frequent pattern association rules is valuable for pattern mining. There are several algorithms for mining positive association rules. Positive means items are together in transactional databases. Positive patterns or item sets form association rules. A positive association rule $A \Rightarrow B$ exists if two item sets, A and B, are purchased together when A is purchased. The existence of several relationships, such as $(\neg A \Rightarrow \neg B)$, $(A \Rightarrow \neg B)$, and $(\neg A \Rightarrow B)$, complicates the search for negative associations. Common issues include discovering frequent and infrequent item sets, suitable positive and negative association rules, single minimum support, and more.

There are many pattern-mining approaches in the literature. Current mining methods include candidate generation methods (partitioning sampling, Apriori, etc.), pattern growth methods (Hmine, FP-growth, Closttt+, FP max, etc.), and vertical format methods. The interesting aspects of mining approaches include subjectivity vs. objectivity, constraint-based mining, mining correlation rules, and exception rules. Database organization further categorizes mining methods: distributed, incremental, and streaming. Database type considerably affects data mining methods. The type of data mined classifies the data mining processes. Time series, sequential, spatial (co-location), structural (lattice, tree, graph), temporal (evolutionary and periodic), video, picture, multimedia, and network patterns can be mined by using methods that include pattern-based clustering, classification, collaborative filtering, semantic annotations, and privacy preservation.

The main problem in finding negative associations is handling huge databases and the right important negative associations. There is a need to consider prominent closed and minimal item sets, especially in the medical field, so focused investigations and corrections can be carried out. Closed frequent item sets limit the patterns generated in frequent item set mining while keeping all information about the set. The set of closed item sets can yield the frequent item sets and their support values. Mining closed item sets is better than all frequent item sets.

Some negative association mining methods have been presented in the literature. No method has been presented that aims at mining frequent, regular, closed, and maximal item sets that are negatively associated with and affected by 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.

The negation of item set A is represented as $\neg A$. The support for $\neg A$ can be computed by subtracting the support of item set A from 1. A rule of the form $(A \Rightarrow B)$ is a positive rule, and other forms of the rules $(A \Rightarrow \neg B)$, $(\neg A \Rightarrow B)$ and $(\neg A \Rightarrow \neg B)$ are negative rules. The confidence of the rules is expressed as $\text{sup}(A \cup \neg B) / \text{sup}(A)$, which measures the interestingness of the negative associations. The negative association rules of type $A \Rightarrow \neg B$ must be discovered, considering that A and B are disjointed and require minimum support and confidence.

$$\text{Support}(i001, \sim i002, i003) = \text{support}(i001, i003) - \text{sup}(i001, i002, i003)$$

$$\text{—supp}(A) \geq \text{ms}, \text{supp}(B) \geq \text{ms}, \text{and } \text{supp}(A \cup B) < \text{ms}, \text{ where ms} = \text{Minimum Support}$$

$$\text{—supp}(A \Rightarrow \neg B) = \text{supp}(A \cup \neg B); \text{ and } \text{—conf}(A \Rightarrow \neg B) = \text{supp}(A \cup \neg B) / \text{supp}(A) \geq \text{mc}, \text{ where mc} = \text{minimum confidence}.$$

Important definitions related to this paper are placed in Appendix A.

1.1. Problem Definition

In the medical field, negative associations are very dangerous to human beings. Incorrect administration of drugs will create many ill health situations rather than cure the real problem. It is necessary to find the contradicting drugs before the same are administered. The frequent and regular drugs, which are numerous, need special attention. There is a need to consider the closedness and maximality of the item set so that fewer frequent and regular item sets can be found and negative associations are determined. The problem is to find the negative associations among the drugs administered due to diagnosing a disease such as CORONA. The use of Beta Locker and Remdesivir to cure CORONA has a direct effect on the heart, leading to a heart attack. The following objectives are set to solve the problem.

1.2. Objectives of the Research

1. To construct a database and generate an example set which can be used to experiment to find the negative associations 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.

2. Related Work

Ming-Syan Chen [6] conducted a detailed assessment of mining processes and their uses. A purpose-based classification of mining methods has been presented. A basic method mines too many patterns, which could result in too many association rules that consumers may find more intriguing. Ashok Savarese [7] provided a mining strategy that uses positive associations and domain expertise to find few negative correlations, making it easy to analyze and present.

Padmanabhan Balaji [8] has shown that pattern mining produces too many patterns and requires decision-makers domain knowledge. Decision makers know precept and belief data. They also used ideas and perceptions to find surprising patterns. They tested WEB log data and found that user perception might be used for efficient mining. Many literature-based mining methods use Apriori-like candidate generation. When dealing with long patterns, this method is time-consuming and expensive.

Jiawe Han [9] proposed combining database compression, pattern fragment growth, and divide and conquer to break mining tasks into small assignments that could be mined under limitations. Their three strategies drastically minimize search space. Pattern mining can be horizontal or vertical. Compared to horizontal mining, vertical mining is effective. Vertical format approaches have the advantage of not requiring fast frequency counting through intersecting transaction IDs and pruning. However, these solutions use more capacity for lighter vertical format table entries. Mohammed J. Zaki [10] introduced Di-set, a vertical data presentation that compares candidate pattern transactions to patterns. They demonstrated how Di-sets can significantly reduce vertical table entry memory.

Many transactional database designs can yield positive and negative association rules. Xindong Wu [11] suggested constructing negative and positive association rules. Negative relationships between patterns were assessed using terms such as $(A \Rightarrow \neg B)$, $(\neg A \Rightarrow B)$, and $(\neg A \Rightarrow \neg B)$. Constraining patterns with interesting patterns mines rules from a vast database. The algorithm finds rules in the form of $\neg X \Rightarrow Y$, $X \Rightarrow \neg Y$, and $\neg X \Rightarrow \neg Y$. With the support confidence framework, the authors included “mininterest”. The dependency between two item sets was checked using mininterest.

Some created association rules may be remarkable due to low interest or high confidence. The Daly et al. [12] technique evaluated exceptional mining rules. They examined

exceptional and negative association rules. Negative association rules are used to generate exception rules. They have also developed a new metric for unusual rule interest. Candidate rules employ extraordinary rules that meet exceptional metrics to examine patterns and decisions.

Most literature-based strategies trim the most desired decision-making patterns using interesting criteria. However, determining the interest measure is difficult and may need trial and error. No accurate approach exists for determining interesting measures. Thiruvady 2004 [13] proposed an approach that uses user inputs to determine the needed rules and constraints. The most intriguing rules are found by using the GRD algorithm.

Statistical correlations determine how effectively two data sets are connected. Maria-Luiza and Antonie [14] developed a method to find negative association rules based on the correlation between two item sets. Negative rules are retrieved if the correlation between item sets is negative and the confidence is high. Chris Cornelis [15] surveyed numerous algorithms that mine negative and positive association rules and outlined several circumstances where the methods presented in the literature failed. They classified and cataloged numerous mining algorithms based on these factors and identified their limitations. They also introduced a modified Apriori mining technique that can detect negative correlations with interesting ones using a confidence framework. They employed an upward closure property that matches validity definitions' support-based interest of negative connections. The interesting parameter "Support" is usually defined for the dataset entry. A hierarchy of data records with support values at each level is acknowledged. Multiple support values are defined at each level. The authors introduced an Apriori-based algorithm (PNAR) that finds negative association rules via upward closure. If $\neg X$ meets the minimal support criterion, then $Y \subseteq I$, $X \cap Y = \emptyset$, and $\neg(XY)$ also do.

MLMS (multi-level minimum support) by Xiangjun Dong [16] defines minimal support values at each record level. MLMS finds frequent and infrequent items. They presented a measure to mine common and infrequent item sets in addition to correlation and confidence. The PNAR-MLMS method generates positively and negatively linked patterns from frequent and infrequent item sets created by the MNMS model. Xiangjun Dong et al. [17] also created PNAR-based classifiers employing association rules divided into recognized categories. Classifiers can then determine if a pattern is negative or positive. Finding the K-most intersecting rule requires minimal support and thresholds. Since the user needs the support value, defining the minimal threshold hold value is problematic. Instead, users can determine interest and the number of rules they expect from the mining algorithm.

The algorithm GRD, which has no minimum support value, is also described in the literature. The user must define the interest and the number of rules they want. Xiangjun Dong [18] extended the GRD approach for mining positive and negative rules. Transactions reveal positive and negative association rules. Negative association principles show how one pattern cancels another. Xiangjun Dong et al. [19] expanded the support confidence framework by adding a sliding correlation coefficient criterion when data availability changes. Correlation coefficients can be determined using several patterns. Antecedents and consequences are positively and negatively connected.

Regular item sets were initially studied by Tanbeer et al. [20]. They proposed a "Regular pattern tree" to find regular patterns. The algorithm scans the database twice. In the first set, item sets' regularity and support values are established, and in the second scan, a regular pattern tree is created. Their method is cyclic and periodic. The minimum support threshold set is used to mine consecutive patterns Weimin Ouyang et al. [21]. The minimum support threshold assumes all common sequences with the same frequency, which is false. Rare item problems occur when pattern sequence frequencies vary despite meeting the minimal threshold value.

Within recurrent patterns, mining negative linkages is just as significant as positive correlations. Indheba Mohammad Ali [22] has developed techniques to mine intriguing negative and positive transactional data connections. Interesting negative and positive association rules (PNAR) and mining interesting multiple-level support methods have been

proposed. Their technique uses different support values to mine positive and negative association rules from intriguing frequent and infrequent item sets. Pavan et al. [23,24] used vertical table mining to uncover positive and negative connections based on item set regularity.

Yanqing Ji et al. [25] focused on mining item sets with casual relationships, which help prevent or correct negative outcomes 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). Their mining algorithm considered the database connecting drugs and 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 [27–30], but none in big data. They have 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 requires more search space than positive 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. [31–33], suggested a Chi-square test for negative association rules. Positive and negative associations were determined using a correlation matrix. 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 item set generated association rules. This domain-specific technique requires a predetermined taxonomy, making it hard to generalize. Similar methods have been presented.

One subclass of negative association rules is found using Teng et al [34,35] substitution rule mining (SRM). The $X \Rightarrow \neg Y$ algorithm identifies negative association rules. First, this algorithm detects “concrete” elements. Concrete things outperform anticipated support with a high Chi-square. The correlation coefficient is determined for each pair.

Using Pearson’s ϕ correlation coefficient, Antonie and Zaiane [36] identified significant positive and negative connection rules. ϕ In GRD, Tiruvady, and Webb’s [37] algorithm, the correlation coefficient for the $X \Rightarrow Y$ connection rule identifies top-k positive and negative associations. More rules can be uncovered using leverage.

Md Saiful Islam et al. [38] conducted a PRISMA-compliant systematic study of healthcare analytics employing data mining and big data. All 2005–2016 articles were reviewed. They ignored unfavorable associations in their review. Hnin et al. [39] have used the maximal frequent item set algorithm for mining item sets from a healthcare database, relevant to heart diseases. A precision 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 not been addressed yet.

Simarjeet Kauri et al. [40] conducted a review using AI techniques to diagnose the disease. No review, however, has been conducted on the drugs administered and the impact of the same when heart diseases are predicted. Jianxiang Wei et al. [41] 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 when negatively associated drugs are administered to a patient.

Lu Yuwen, Shuyu et al. [42] proposed a framework that uses sequential data mining called “Prefix-Span” and a disproportionality-based method called “Proportional Re-

port Ratio” to detect serious adverse drug reactions based on casual relationships, drugs, and drug reactions. They examined single drug-to-drug responses. Constricting medication responses, which are harmful, have not been explored.

Yifeng Lu et al. [43] have presented that frequent item set mining reveals expected patterns. The negative associations between the drugs recommended can thus be found. But the issue is that an infrequent item set, which also has negative associations among the drugs, is also crucial. The authors have presented a method for mining infrequent closed item sets using bi-directional traversing. However, the negative associations among infrequent or frequent item sets have yet to be explored.

Jingzhuo Zhang et al. [44] 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. A bidirectional encoder representation from transformers has been used to extract the relationships between the drugs. However, no modeling is carried out to classify whether the interactions are positive or negative.

E. Ramaraj et al. [45] 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 focused on negative associations among the drugs based on chemical compositions or diseases based on drug reactions. SPNAR, developed by Chris Cornelis et al. [46], mines positive and negative rules. They also proposed BAECLAT for mining positive and negative rules from large databases. Mario Luiza Anionic et al. [47] have proposed a modified “BAEECLAT” method for mining positive and negative association rules by confining the rules.

Jigar R. Desai et al. [48] 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 diabetes (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. Considering their chemical compositions, the study did not focus on the negative associations among the drugs.

M. Goldacre et al. [49] 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. [50] 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 learned 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.

Table 1 compares algorithms for negative association mining. The table shows that maximality and closure were ignored when determining the negative associations. 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 item set. 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.

Table 1. Comparative analysis of mining algorithms concerning negative association mining.

[illegible]

3. Methods for Computing Negative Association among Frequent, Regular, Closed and Maximal Item Sets

3.1. Method 1

If item sets X and Y are numerous but rarely occur together, then $\text{sup}(X \cup Y) < \text{sup}(X) * \text{sup}(Y)$, indicating a negatively correlated pattern.

If $\text{sup}(X \cup Y) < \text{sup}(X) * \text{sup}(Y)$, X and Y are substantially negatively correlated, resulting in a strongly negatively correlated pattern. This definition can be extended to k -item sets. However, null transactions occur.

3.2. Method 2

$\text{Sup}(X \neg E) * \text{sup}(\neg A \cup E) < \text{sup}(X \cup E) * \text{sup}(\neg A \cup \neg E)$, causing a null transaction to indicate the existence of a negative association.

3.3. Method 3

Suppose that item sets A and B 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, $P(A|B) + P(B|A)/2 < \epsilon$, where ϵ is the negative pattern threshold. This way of computing the negative association is free from the problem of null transactions.

4. Computing the Negative Associations from Regular, Frequent, Closed and Maximal Item Sets

The algorithm that mines negative associations from regular, frequent, closed and maximal item sets is shown in Algorithm 1.

Algorithm 1 Mining negative associations from regular, frequent, closed and maximal item sets

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 data from a flat file/DBMS table into an array, as illustrated in Table 2.
 3. Convert Table 2 data to vertical format as displayed in Table 3.
 4. Prune initial irregular and non-frequent items implies deleting such records from Table 3.
 5. Find closed and maximal item sets and place the same into Table 4.

For every record in Table 3

{

Selecting Closed and Maximal Item Sets
 Suppose the item set is a subset of the existing data set in Table 3. Loop.
If the item set in the record is a superset of any other record in Table 3
 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.
 else
 Add the record to Table 4 as a close maximal item set.

 }
 6. For every record in Table 4

For every next record in Table 4

 - i. Find the intersection of the current and next records.
 - ii. If the intersection is null, enter the current and next items into a negative association Table 5.
 - iii. If the intersection is not null, find the common items.
 1. If the count of elements is $>$ the regularity threshold and the frequency threshold, add the records to Table 4 at the end.
 2. LOOP if the common elements do not satisfy the regularity or frequency constraint.
 7. For each of the negatively associated chemicals shown in Table 5, find the related drugs and report the negative associations among the drugs.
-

Table 2. Sample medical data extracted from the database.

P.SL.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.

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

5. 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. In total, 100,000 patient registrations 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 each 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.

6. Results and Discussion

6.1. Results—Implementation of the Algorithm 1 on the Dataset

Step 1: Extract sample data from the database.

Algorithm 1 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.

Step 2: Add transaction IDs to the extracted data from the database.

Transaction IDs are assigned to the extracted data as shown in Table 2 (Column 2) to keep track of each of the transactions. Table 2 lists the first 23 records of the database. Table 2 contains both the extracted data and the transaction IDs added to the data.

Step 3: Convert Table 2 into a vertical format.

Table 2 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 are computed, and the same are shown in Table 3. Regularity is computed based on the relative positions of the record in the database, and the frequency is computed based on the occurrence count.

Step 4: Prune the records that do not meet the threshold levels of regularity and frequency.

The maximum regularity (4) and the minimum frequency (3), recommended by the users, are 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 regularity and frequency threshold value. The records left over are shown in Table 4. Using this criterion, five chemicals have been pruned.

Step 5: Prune the records which do not 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 5.

Step 6: 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 $(CH4 \Rightarrow CH8)$, $(CH4 \Rightarrow CH11)$, $(CH4 \Rightarrow CH8, CH11)$, $(CH6 \Rightarrow CH8)$, $(CH6 \Rightarrow CH11)$, $(CH6 \Rightarrow CH8, CH11)$, $(CH8 \Rightarrow CH4, CH6)$, $(CH11 \Rightarrow CH4, CH6)$, $(CH4, CH6) \Rightarrow CH8, CH11$.

Step 7: Find negatively associated drugs

Map back 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

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

6.2. Discussion

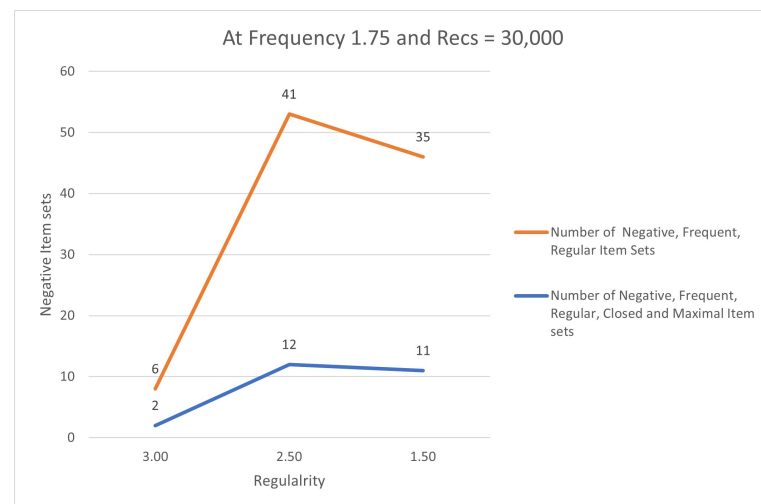
The 100,000 examples have been created through data collection and analyzed for different sample sizes (30,000, 50,000 and 70,000). Different thresholds were fixed concerning regularity and support, and a number of negative associations were found by applying support and regularity thresholds. The records are processed with and without maximality+ closedness applied.

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 mined using Algorithm 1. The negative associations have been generated by keeping the regularity fixed and varying the support.

Table 8 shows the variation in the number of negative associations, fixing the frequency at 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 is reduced by 46.90% when the closedness and maximal criteria are also considered.

Table 7. Analysis of negative frequent regular item set 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

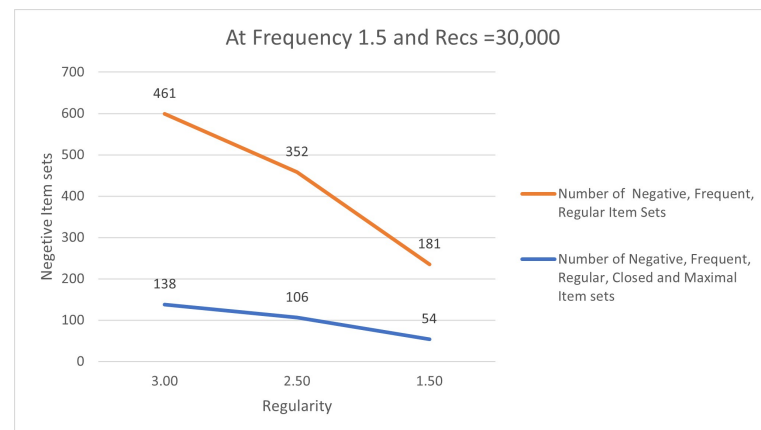
**Figure 1.** Variance of negative association item sets fixing the frequency (1.75) and varying the regularity for 30,000 example records.**Table 8.** Number of negative associations considering two criteria, fixing the frequency at 1.75 and varying regularity when 30,000 examples are selected.

At Frequency 1.75 and Recs = 30,000			
Regularity	Number of Negative Frequent Regular Item Sets	Number Negative Frequent Regular Item, Closed and Maximal Item Sets	% Decrease
3.00	6	2	66.7
2.50	41	12	70.7
2.00	35	11	68.6
Average % of decrease in negative associations			68.6

Table 9 shows the variation in the number of negative associations, fixing the frequency at 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 is reduced by 70.00% 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 Negative Frequent Regular Item Sets	Number Negative Frequent Regular Item, Closed and Maximal Item Sets	% Decrease
3.00	461	138	70.1
2.50	352	106	69.9
2.00	181	54	70.2
Average			70.0

**Figure 2.** Variance of negative association item sets fixing the frequency (1.50) and varying the regularity for 30,000 example records.

Further analysis has been carried out considering the higher size of the example set. Table 10 shows the percentage of negative associations reduction 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%

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 with 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 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

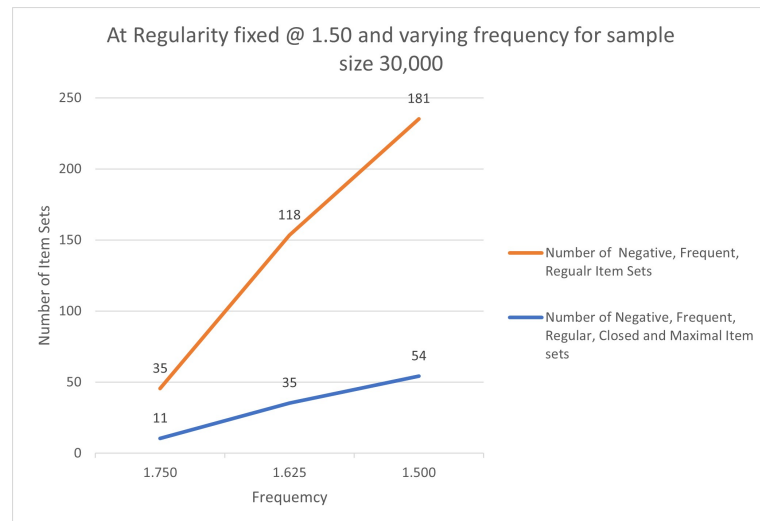


Figure 3. Variance of negative association item sets fixing the regularity (1.50) and varying the frequency for 30,000 example records.

Table 11. Analysis of variations in negative associations fixing the regularity at 1.50 and varying frequency for a 30,000-sample size.

Frequency	Number Negative Frequent Regular Item Sets	Number Negative Frequent Regular Item, Closed and Maximal Item Sets	% Reduction
1.750	35	11	70.0
1.625	118	35	70.0
1.500	181	54	70.0
Average			70.0

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.

Table 12. Number of negative associations fixing regularity at 1.65, no of records at 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

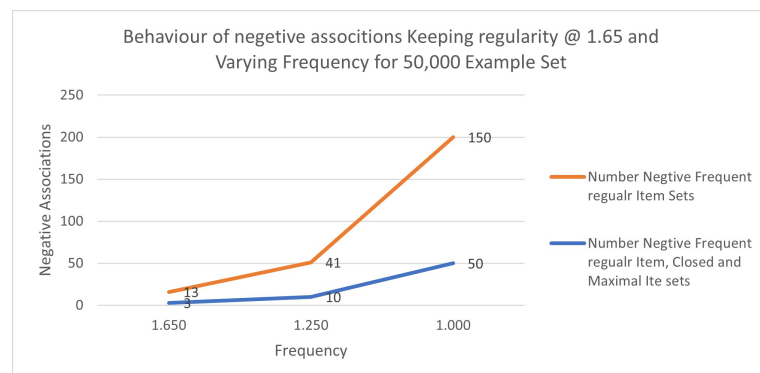


Figure 4. Variance of negative association item sets fixing the regularity (1.65) and varying the frequency for 50,000 example records.

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%

7. Conclusions and Future Scope

Finding negative associations among the drugs administered to patients due to the diseases is extremely important as it saves many lives and provides excellent help to doctors in prescribing proper medicines. Many have suffered during the COVID-19 period due to too many side reactions caused due to administering drugs to cure COVID-19. Remedying drugs, when given with other drugs, created several side reactions such as black fungus, heart attacks and many more.

Finding minimal and most effective negative associations is crucial so that the focus is on critical item sets. Minimal negative associations, the most effective negative associations, can be found through mining item sets that are regular, frequent, maximal, and closed. A reduction of 70% of negative associations has been achieved considering any threshold on frequency and regularity.

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.

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. The most important limitation of this research is the requirement of the chemical composition of the drugs, which doctors generally prescribe to cure diseases.

Further research can be carried out considering the constraints and colossal patterns that should be imposed when administering certain types of drugs. The problem must be investigated, considering rare item sets and new ones added to the medical system.

Further research can also be conducted to find specialized interesting measures different from frequency and support that directly suit to find effective negative associations. The algorithm can be extended for finding negative associations from distributed medical databases, or when incremental medical databases are added, and the medical data are made available in streaming mode.

Author Contributions: Conceptualization, S.K.R.J.; methodology, S.K.R.J.; software, R.R.B.; validation, R.R.B.; formal analysis, S.K.R.J.; investigation, S.K.R.J.; resources, R.R.B.; data curation, R.R.B.; writing—original draft preparation, S.K.R.J.; writing—review and editing, S.K.R.J.; visualization, R.R.B.; supervision, S.K.R.J.; project administration, S.K.R.J.; funding acquisition, R.R.B. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: No ethical approval is required for this study.

Data Availability Statement: The data presented in this study will be made available on request from the corresponding author after seeking approval from the hospitals. The data are not publicly available due to non-disclosures signed with the Hospitals.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A. Important Definitions and Representations

This Appendix table provides important definitions and the related empirical formulations used in formulating the method proposed in the paper.

Key Element	Description of the Key Element
Item Set	A set of Items that together appear in a Transaction
K-Item Set	An item set having k number of items
Closed Item Sets	An Item set is closed in a data set D if no super item set Y exists, such that Y has the same support count as X in D.
Frequency of an item set	The number of transactions containing a specific item set is called absolute support. An item set is said to be frequent if the relative support or absolute support is less than >the minimum threshold value'.
Regular item sets	Regular item sets are those that occur many times within a specific time. The periodicity could be absolute or relative, which is measured as the distance between the transactions containing an item set. The regularity of the item set plays a major role when an unexpected disease occurs every time a specific drug is administered.
Closed frequent item sets	An item set X is a closed frequent item set in D if X is both closed and frequent in D.
Maximal frequent item set	An item set X is the maximal frequent item set in D when X is frequent, and no super item set Y exists such that $X \subset Y$ and Y are frequent in D.
Closed and Maximal Item set.	An item set is closed and maximal if the principality of maximality applies to closed item sets. Closed and maximal 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 frequent item sets rather than all set of all frequent item sets.
Interestingness measures	A set of measures (support, confidence, correlation, etc.) that reveal the interestingness of an item set for the user
Association Rule	Is a rule applied to a set of item sets and triggers whether a rule reveals a positive association or negative association among the item sets
Support of an association rule	The support of the association rule $A \rightarrow B$ is the percentage of transactions that contain $A \cup B$ or $P(A \cup B)$
Confidence of an association rule	The confidence of the rule $A \rightarrow B$ is the percentage of transactions in D that contain A and B. This is taken to be the conditional probability $P(B A)$
Strong association rules	The rules that satisfy minimum support and minimum confidence are called strong association rules.
Correlation	Item sets A and B are set to be correlated if the correlation coefficient between A and B is positive. Lift is a measure of the correlation between the item sets, and item sets A and B are independent when
Lift	$= P(B A) / P(B) = \text{conf}(A \Rightarrow B) / \text{sup}(B)$
Item merging	It is a pruning method to reduce the number of items for finding the applicable rules. If every transaction containing a frequent item set X also contains an item set Y and X is not a superset of Y, then $X \cup Y$ forms a closed item set, and there is no need for searching for item set X but no Y.
Sub item pruning	It is a pruning method to reduce the number of items for finding the applicable rules. If a frequent item X is a proper subset of an already found frequent closed item set Y and support count (X) = support count (Y), then X and all its descendants in the set enumeration tree cannot be frequent closed item sets and thus can be pruned.
Item skipping	It is a pruning technique when a database is mined to find a hierarchical structure by employing depth-first; mining closed item sets at each level is undertaken. An item set X is associated with a header table and projected database. If a local frequent item p has the same support in several header tables at different levels, prune p from the header table at higher levels.

References

- Aggarwal, C.C.; Yu, P.S. Mining associations with the collective strength approach. *IEEE Trans. Knowl. Data Eng.* **2001**, *13*, 863–873. [\[CrossRef\]](#)
- Aggarwal, C.C.; Yu, P.S. A new framework for item-set generation. In Proceedings of the Seventeenth ACM SIGACTSIGMOD-SIGART Symposium on Principles of Database Systems, PODS'98, Seattle, WA, USA, 1–4 June 1998; pp. 18–24.
- Agrawal, R.; Imielinski, T.; Swami, A. Mining association rules between sets of items in large databases. In *ACM SIGMOD Record*; ACM Press: New York, NY, USA, 1993; pp. 207–216.
- Agrawal, R.; Srikant, R. Fast algorithms for mining association rules. In Proceedings of the VLDB 1994 Proceedings of the 20th International Conference on Very Large Data Bases, San Francisco, CA, USA, 12–15 September 1994; pp. 487–499.
- Mahmood, S.; Shahbaz, M.; Guergachi, A. Negative and positive association rules mining from text using frequent and infrequent item sets. *Sci. World J.* **2014**, *2014*, 973750. [\[CrossRef\]](#) [\[PubMed\]](#)

6. Chen, M.; Han, J.; Yu, P.S. Datamining an overview from database perspective. *IEEE Trans. Knowl. Data Eng.* **1996**, *8*, 866–883. [\[CrossRef\]](#)
7. Ashok Savasere, A.; Omiecinski, E.; Navathe, S. Mining for strong negative associations in a large database of customer transactions. In Proceedings of the Fourteenth International Conference on Data Engineering, Orlando, FL, USA, 23–27 February 1998; pp. 494–502.
8. Padmanabhan, B.; Tuzhilin, A. A belief-driven method for discovering unexpected patterns. In Proceedings of the Fourth International Conference on Knowledge Discovery and Data Mining (KDD-98), New York, NY, USA, 27–31 August 1998; pp. 94–100.
9. Han, J.; Yin, Y. Mining Frequent Patterns without candidate generation. *ACM SIGMOD Rec.* **2000**, *29*, 1–12. [\[CrossRef\]](#)
10. Zaki, M.J. Fast Vertical Mining Using Diffsets. In Proceedings of the KDD03: The Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Washington, DC, USA, 24–27 August 2003.
11. Wu, X.; Zhang, C.; Zhang, S. Efficient Mining of Positive and Negative Association Rules. *ACM Trans. Inf. Syst.* **2004**, *22*, 381–405. [\[CrossRef\]](#)
12. Daly, O.; Taniar, D. Exception Rules Mining Based On Negative Association Rules. *Lect. Notes Comput. Sci.* **2004**, *3046*, 543–552. [\[CrossRef\]](#)
13. Thiruvady, D.R.; Webb, G.I. Mining Negative Association Rules Using GRD. In Proceedings of the Pacific-Asia Conference on Knowledge Discovery and Data Mining, Sydney, Australia, 26–28 May 2004; pp. 161–165. [\[CrossRef\]](#)
14. Antonie, M.; Zaiane, O.R. Mining Positive and Negative Association Rules: An Approach for Confined Rules. European Conference on Principles of Data Mining and Knowledge Discovery. Mining Positive and Negative Association Rules An Approach for Confined Rules. In Proceedings of the 8th European Conference on Principles and Practice of Knowledge Discovery in Databases (PKDD04), Pisa, Italy, 20–24 September 2004; pp. 27–38. [\[CrossRef\]](#)
15. Cornelis, C.; Yan, P.; Zhang, X.; Chen, G. Mining Positive and Negative Association Rules from Large Databases. In Proceedings of the 2006 IEEE Conference on Cybernetics and Intelligent Systems, Bangkok, Thailand, 7–9 June 2006. [\[CrossRef\]](#)
16. Dong, X.; Sun, F.; Han, X.; Hou, R. Study of Positive and Negative Association Rules Based on multi-confidence and Chi-Squared Test. In *Advanced Data Mining and Applications*; Li, X., Zaiane, O.R., Li, Z., Eds.; Springer: Berlin/Heidelberg, Germany, 2006; Volume 4093, pp. 100–109. [\[CrossRef\]](#)
17. Dong, X.; Niu, Z.; Shi, X.; Zhang, X.; Zhu, D. Mining Both Positive and Negative Association Rules from Frequent and Infrequent Itemsets. In Proceedings of the Third International Conference on Advanced Data Mining and Applications (ADMA 2007), Harbin, China, 6–8 August 2007; pp. 122–133. [\[CrossRef\]](#)
18. Dong, X.; Zheng, Z.; Niu, Z.; Jia, Q. Mining Infrequent Item sets based on Multiple Level Minimum Supports. In Proceedings of the Second International Conference on Innovative Computing, Information and Control, Kumamoto, Japan, 5–7 September 2007.
19. Dong, X.; Niu, Z.; Zhu, D.; Zheng, Z.; Jia, Q. Mining Interesting Infrequent and Frequent Itemsets Based on MLMS Model. In Proceedings of the International Conference on Advanced Data Mining and Applications, Chengdu, China, 8–10 October 2008; pp. 444–451.
20. Khairuzzaman, T.S.; Ahmed, C.F.; Jeong, B.; Lee, Y. Mining regular patterns in transactional databases. *IEICE Trans. Inf. Syst.* **2008**, *91*, 2568–2577.
21. Ouyang, W.; Huang, Q. Mining Positive and Negative Sequential Patterns with Multiple Minimum Supports in Large Transaction Databases. In Proceedings of the 2010 Second WRI Global Congress on Intelligent Systems, Wuhan, China, 16–17 December 2010. [\[CrossRef\]](#)
22. Swesi, I.M.A.O.; Bakar, A.A.; Kadir, A.S.A. Mining Positive and Negative Association Rules from Interesting Frequent and Infrequent Itemsets. In Proceedings of the 2012 9th International Conference on Fuzzy Systems and Knowledge Discovery, Chongqing, China, 29–31 May 2012.
23. Kumar, N.V.S.P.; Rao, K.R. Mining Positive and Negative Regular Item-Sets using Vertical Databases. *Int. J. Simul. Syst. Sci. Technol.* **2016**, *17*, 33.1–33.4. [\[CrossRef\]](#)
24. Kumar, N.V.S.P.; Rao, L.J.J.; Kumar, G.V. A Study on Positive and Negative Association rule mining. *Int. J. Eng. Res. Technol. (IJERT)* **2012**, 1–4.
25. Ji, Y.; Ying, H.; Tran, J.; Dews, P.; Mansour, A.; Massanari, R.M. A Method for Mining Infrequent Causal Associations and Its Application in Finding Adverse Drug Reaction Signal Pairs. *IEEE Trans. Knowl. Data Eng.* **2013**, *25*, 721–733. [\[CrossRef\]](#)
26. Bagui, S.; Dhar, P.C. Positive and negative association rule mining in Hadoop’s MapReduce environment. *J. Big Data* **2019**, *6*, 75. [\[CrossRef\]](#)
27. Jiang, H.; Luan, X.; Dong, X. Mining weighted negative association rules from infrequent item sets based on multiple support. In Proceedings of the 2012 International Conference on Industrial Control and Electronics Engineering, Washington, DC, USA, 23–25 August 2012; pp. 89–92.
28. Kishor, P.; Porika, S. An efficient approach for mining positive and negative association rules from large transactional databases. In Proceedings of the 2016 International Conference on Inventive Computation Technologies (ICICT), Coimbatore, India, 26–27 August 2016.
29. Ramasubbareddy, B.; Govardhan, A.; Ramamohanreddy, A. Mining positive and negative association rules. In Proceedings of the 5th International Conference on Computer Science and Education, Hefei, China, 24–27 August 2018; pp. 1403–1406.

30. Sahu, A.K.; Kumar, R.; Rahim, N. Mining negative association rules in a distributed environment. In Proceedings of the 2015 International Conference on Computational Intelligence and Communication Networks, Jabalpur, India, 12–14 December 2015; pp. 934–937.
31. Brin, S.; Motwani, R.; Silverstein, C. Beyond market basket: Generalizing association rules to correlations. In Proceedings of the 1997 ACM SIGMOD International Conference on Management of Data ACM SIGMOD, Tucson, AZ, USA, 13–15 May 1997; pp. 265–276.
32. Antonie, L.; Li, J.; Zaiane, O. Negative association rules. In *Frequent Pattern Mining*; Springer: Cham, Switzerland, 2014; pp. 135–145.
33. Savasere, A.; Omiecinski, E.; Navathe, S. Mining for strong negative associations in a large database of customer transactions. In Proceedings of the ICDE, Orlando, FL, USA, 23–27 February 1998; pp. 494–502.
34. Teng, W.-G.; Hsieh, M.-J.; Chen, M.-S. On the mining of substitution rules for statistically dependent items. In Proceedings of the ICDM, Maebashi City, Japan, 9–12 December 2002; pp. 442–449.
35. Teng, W.G.; Hsieh, M.-J.; Chen, M.-S. A statistical framework for mining substitution rules. *Knowl. Inf. Syst.* **2005**, *7*, 158–178. [\[CrossRef\]](#)
36. Antonie, M.-L.; Zaiane, O.R. *Mining Positive and Negative Association Rules: An Approach for Confined Rules*; Lecture Notes in Computer Science; Springer: Berlin/Heidelberg, Germany, 2004; pp. 27–38.
37. Thiruvady, D.R.; Webb, G.I. Mining negative rules using GRD. In *Advances in Knowledge Discovery and Data Mining*; Lecture Notes in Computer Science; Springer: Berlin/Heidelberg, Germany, 2004; Volume 3056, pp. 161–165.
38. Islam, M.S.; Hasan, M.M.; Wang, X.; Germack, H.D.; Noor-E-Alam, M. A Systematic Review on Healthcare Analytics: Application and Theoretical Perspective of Data Mining. *Healthcare* **2018**, *6*, 54. [\[CrossRef\]](#) [\[PubMed\]](#)
39. Khaing, H.W. Data Mining based Fragmentation and Prediction of Medical Data. In Proceedings of the 2011 3rd International Conference on Computer Research and Development, Shanghai, China, 11–13 March 2011; pp. 480–485.
40. Kaur, S.; Singla, J.; Nkenyereye, L.; Jha, S.; Prashar, D.; Joshi, G.P.; El-Sappagh, S.; Islam, M.S.; Islam, S.M. Medical Diagnostic Systems Using Artificial Intelligence (AI) Algorithms: Principles and Perspectives. *IEEE Access* **2020**, *8*, 228049–228069. [\[CrossRef\]](#)
41. Wei, J.; Lu, Z.; Qiu, K.; Li, P.; Sun, A.H. Predicting Drug Risk Level from Adverse Drug Reactions Using SMOTE and Machine Learning Approaches. *IEEE Access* **2020**, *8*, 185761–185775. [\[CrossRef\]](#)
42. Lu, Y.; Chen, S.; Zhang, H. Detecting Potential Serious Adverse Drug Reactions using Sequential Pattern Mining Method. In Proceedings of the 2018 IEEE 9th International Conference on Software Engineering and Service Science (ICSESS), Beijing, China, 23–25 November 2018; pp. 56–59.
43. Lu, Y.; Seidl, T. Towards Efficient, Closed Infrequent Item set Mining using Bi-directional Traversing. In Proceedings of the 2018 IEEE 5th International Conference on Data Science and Advanced Analytics (DSAA), Turin, Italy, 1–3 October 2018; pp. 140–149.
44. Zhang, J.; Liu, W.; Wang, P. Drug-Drug Interaction Extraction from Chinese Biomedical Literature, using distant supervision. In Proceedings of the 2020 IEEE International Conference on Knowledge Graph (ICKG), Nanjing, China, 9–11 August 2020; pp. 593–598.
45. Ramaraj, E.; Venkatesan, N. Positive and Negative Association Rule Analysis in Health Care Database. *IJCSNS Int. J. Comput. Sci. Netw. Secur.* **2008**, *8*, 325–330.
46. Antonie, M.L.; Zaiane, O.R. Mining Positive and Negative Association Rules: An Approach for Confined Rules. In Proceedings of the 8th European Conference on Principles and Practice of Knowledge Discovery in Databases, Pisa, Italy, 20–24 September 2004; pp. 27–38.
47. Antonie, M.L.; Zaiane, O.R. Mining Positive and Negative Association Rules—An approach for confine rules. In *European Conference on Principles of Knowledge Discovery*; Springer: Berlin/Heidelberg, Germany, 2004; pp. 27–38.
48. Desai, J.R.; Hyde, C.L.; Kabadi, S.; Louis, M.S.; Bonato, V.; Loomis, A.K.; Galaznik, A.; Berger, M.L. Utilization of Positive and Negative Controls to Examine Comorbid Associations in Observational Database Studies. *Med. Care* **2017**, *55*, 244–251. [\[CrossRef\]](#)
49. Goldacre, M.; Kurina, L.; Yeates, D.; Seagroatt, V.; Gill, L. Use large medical databases to study disease associations. *QJM Int. J. Med.* **2000**, *93*, 669–675. [\[CrossRef\]](#) [\[PubMed\]](#)
50. Kim, Y.; Cho, Y. Predicting Drug–Gene–Disease Associations by Tensor Decomposition for Network-Based Computational drug repositioning. *Biomedicines* **2023**, *11*, 1998. [\[CrossRef\]](#) [\[PubMed\]](#)

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.