Contemporary Mathematics
https://ojs.wiserpub.com/index.php/CM/

Research Article

Finding Negative Associations from Medical Data Streams Based on Frequent and Regular Patterns

Sastry Kodanda Rama Jammalamadaka[1*], Raja Rao Budaraju[2]

[1] Department of Electronics and Computer Science, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Guntur, 522302, Andhra Pradesh, India
[2] Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Guntur, 522302, Andhra Pradesh, India
E-mail: drsastry@kluniversity.in

Received: 10 December 2024; **Revised:** 5 February 2025; **Accepted:** 10 February 2025

Abstract: Medical data flows in streams as the data related to clinical tests and administered drugs by the doctors flows continuously. The doctors must be immediately alerted if negative associations are found among the drugs they prescribe. Data streams are to be processed in single scans as it is not possible to rescan the data for any iterative processing. To detect negative drug connections, regular and frequent drug patterns must be processed. Negative correlations between disease-curing medications might create adverse responses that kill patients. This paper proposes an algorithm that finds the negative associations among regular and frequent patterns mined from medical data streams. The algorithm mines the most effective and critical negative associations. By enforcing optimum frequency and regularity, the number of negative associations reduced to 0.43 from 0.73 for 1,000 item sets mined.

Keywords: data streams, negative associations, adverse effects, side reactions, frequent, regular patterns

MSC: 65L05, 34K06, 34K28

1. Introduction

1.1 *Introduction to data streams*

Data streams are sequences of data elements made available over time. These elements include integers, characters, JavaScript Object Notation (JSON) objects, or Extensible Markup Language (XML) documents. Computer science, telecommunications, and data analysis use data streams. Data streams can be limitless. Finite datasets have all data upfront, whereas data streams can last forever [1]. Traditional techniques for static datasets may not work for this, making processing and analysis difficult. Data streams are sequences of data elements made available over time. These elements include integers, characters, JSON objects, or XML documents. Computer science, telecommunications, and data analysis use data streams [2].

Due to novel applications like monitoring, supply chain execution, sensor networks, oilfield and pipeline operations, financial marketing, and health data, data streams have garnered attention in data analysis and mining. Telecommunications have made stream data from numerous applications easier to access. Static data in data warehouses or databases differs

DOI: https://doi.org/10.37256/cm.6220256229

from stream data. Continuous, high-speed data streams vary over time. Traditional data mining algorithms presume data is stored centrally and can be processed several times, utilising powerful processors to provide offline output without a time limit [3]. These algorithms are unsuitable for dynamic data streams. Stream data may not be stored; thus, it must be mined quickly. Streams also reflect the environment generating them so that early analysis can uncover problems, delays, performance assessments, trend analysis, and other diagnostics. Association rules have to be mined to find the negative associations. The association rules can be incrementally mined using the data available centrally. Association rules with a lift value larger than 1 and a minimal confidence criterion are interesting. Strong association rules exist.

1.2 *Introduction to negative associations*

Support is a measure frequency of an itemset in a dataset. It is measured as per Equation (1)

$$\textbf{Support }(\mathbf{X}) = \frac{\textbf{Number of Transactions containing X}}{\textbf{Total Number of Transactions}} \tag{1}$$

Support helps to determine the significance of an itemset. Higher support means the itemset is more common. Support can be used to filter out infrequent or irrelevant patterns.

Confidence measures the likelihood that if a transaction contains itemset **X**, it will also contain itemset **Y**. It is used in **association rule mining** to evaluate the strength of a rule. It is measured as per Equation (2)

$$\textbf{Confidence }(\mathbf{X} \Rightarrow \mathbf{Y}) = \frac{\textbf{Support }(\mathbf{X\ U\ Y})}{\textbf{Support }(\mathbf{X})} \tag{2}$$

It expresses how often **Y** appears when **X** is present. A higher confidence value indicates a stronger association.

The negation of an item set A is $\neg A$. To calculate the support of $\neg A$, subtract the support of itemset A from 1. ($A \Rightarrow B$) is a positive rule, while ($\neg A \Rightarrow B$), ($A \Rightarrow \neg B$), and ($\neg A \Rightarrow \neg B$) are negative rules. The confidence of rules is measured by the interestingness of negative associations (sup ($A \cup \neg B$)/sup (A).

The negative association rules of type $A \Rightarrow \neg B$ must be identified, when A and B are disconnected and require minimal support and confidence. The implications of negative associations are shown in the Equations (3)-(6)

$$\text{Support (i001, \textasciitilde i002, i003)} = \text{support (i001, i003)} - \text{sup (i001, i002, i003)} \tag{3}$$

Support (i001, ~i002, i003) implies existence of i001, i003 and nonexistence of i002

$$\text{supp } (A) \geq \text{ms, supp } (B) \geq \text{ms and supp } (A \cup B) < \text{ms; where ms} = \text{Minimum Support} \tag{4}$$

$$\text{supp } (A \Rightarrow \neg B) = \text{supp } (A \cup \neg B). \tag{5}$$

Support of (A implies $\neg B$) is same as supp ($A \cup \neg B$) through the implications rule

$$\text{conf } (A \Rightarrow \neg B) = \text{supp } (A \cup \neg B)/\text{supp}(A) \geq \text{mc where mc} = \text{Minimum Confidence.} \tag{6}$$

1.3 *The conf $(A \Rightarrow \neg B)$ implies supp $(A \cup \neg B)$/supp(A) through use of implication rule*

Negatively linked patterns conflict. When confronted with a medicine with a different chemical composition or weather forecast, patterns conflict, leading to inaccurate conclusions and actions. Negative relationships can be more important than positive ones and demand a clear investigation.

Negative association laws say negatively related goods decline if one rises. Negative association rule mining can help improve healthcare crime data analysis decision support systems [4]. The influence negative patterns have on a healthcare system is more drastic and needs to be addressed. Finance, medicine, and prediction have negative trends. Two drugs with different components may clash. Cool zones may not follow temperature rules. Finding frequent, regular, and negative trends is important as they have the most drastic effects on the systems. Negative signifies absence. Two or more item sets may conflict which also are dubbed nonoverlapping patterns.

Even if support is below the threshold, negative association rules develop when two item sets have a negative correlation and high confidence. Some typical item sets must be together. Patterns with a negative correlation between item sets indicate that one set does not affect the other. An item set "A" is negatively associated with the item set "B" when only one of them occurs. Unique patterns with negative words may be rules. Unexpected patterns are association rule exceptions. Negative frequent pattern relationships help pattern mining. Finding negative associations is complicated by the presence of several linkages, such as $(\neg A \Rightarrow \neg B)$, $(A \Rightarrow \neg B)$, and $(\neg A \Rightarrow B)$. Common challenges include frequent and infrequent item sets, negative association rules, minimum support, maximum regulations, etc.

Most medical data is generated continuously and flows in streams. Doctors carry diagnoses and prognoses and continuously prescribe medicines. Doctors need to be altered instantly when negative associations are found among the drugs they prescribe. Association rule mining is a prominent data mining method for finding relationships between the items which are either positive or negative [5–7]. Most of the medical data to be processed is huge. The popular association rule mining method is Apriori, which consumes too much time and is generally not suitable for mining data streams [8].

Many modern applications require capturing continuous data generated by man-held wearable devices. The data flowing through those streams must be processed, and any negative associations must be reported to the people concerned. Since data is detected and transported quickly, instream data which is huge must be processed instantly. Unlike databases, since instream data is never kept raw, it can only be scanned once.

Sensors and other streaming devices collect data and transmitted instantly. The processed findings are briefly kept for analysis and decision-making. The instream data must be processed sequentially to uncover regular and frequent patterns with negative relationships.

Positive association rule mining is used for online log data, census data, biological data, fraud detection, and more. The number of negative associations decreases as the frequency of the item set increases. A basic technique mines too many patterns, creating too many association rules in which consumers may not be interested [9].

Di-Sets helps a vertical data presentation, which helps in comparing candidate patterns and reduces vertical table entry memory [10].

Yang et al. [11] have presented that Frequent Pattern (FP)-Tree and Rare Pattern (RP)-Tree are more stable and effective for mining positive patterns. Several other methods were developed to find sequential, irregular, unusual patterns. Little focus is on uncovering rare negative patterns. Real-time settings are important for finding negative patterns, such as penguins being birds but not flying. Patterns are item sets that occur in multiple transactions. Processing enormous numbers of patterns will take time. As the database grows, so do patterns. Frequent patterns matter more. Selecting patterns that fulfil the user-specified minimum threshold value finds frequent patterns. The user-selected support value is intriguing. The pattern frequency is how often it appears in the database. No timeframe is set here. Sporadic frequent patterns may not have a regular occurrence behavior. Most of these algorithms focus on mining positive patterns and do not focus on a specific domain, and also the negative associations.

Gulzar et al. [12] have presented "An Efficient Healthcare Data Mining Approach Using Apriori Algorithm and used the same to study Eye Disorders in Young Adults". They have not used the healthcare data that flows in streams. The appropriate method as such is time-consuming. They have not addressed the issue of negative associations.

Regular patterns can be inconsistent with frequent ones. Sometimes, both must be considered. Association between mined patterns is crucial. Positive connections between regular or regular-frequent, regular-frequent-maximal-closed patterns are commonly considered.

1.4 *Mining negative associations from data streams*

Data streams can be limitless. Finite datasets have all data upfront, whereas data streams can last forever. Finding novel ways to mine negative associations from continuously flowing data streams is difficult. Windowing, Sampling, and Summarization are used in data streaming applications to analyse and extract insights from the continuous flow of data; however, negative associations in streamed data have not been addressed.

The data to be handled greatly affects pattern finding. Streaming data must be treated as flowing, not stored. Data flows are recognised and acted upon. Dynamic data streams have complex variables and objects. Negative connections dominated the investigation at a minimum frequency and maximum regularity. Negative connections from continuous data streams are to be mined with little significance. Identifying negative connections in medical data from multiple sources and alerting clinicians to their threats through a challenge.

1.5 *Problem definition*

Most of the methods presented in the literature concentrated on mining positive associations in data streams; they did not focus on mining negative associations from data streams. No research concentrated on medical data that flowed in streams and attempted to mine negative associations.

Negative associations among chemicals of various drugs are dangerous as they lead to adverse effects on the patients. Medical data moves in streams continuously from the hospitals, which needs to be processed soon after the data is captured, and the negative associations are to be found and reported immediately. The frequent and regular items that meet threshold criteria must be found and see if the item sets form any negative associations.

The problem is to mine the negative associations in the data streams containing the data relating to the drugs prescribed after diagnosing a disease such as CORONA.

1.6 *Motivation*

The need to find the solutions to this problem has been selected during the CORONA times and in recent times when many have died due to brain tumours. The use of Beta Locker and Remdesivir to cure CORONA has a direct effect on the heart, leading to a heart attack. Using chemotherapy to solve the brain tumours attacks the Human heart. Such negative associations need to be found to help doctors correct their prescriptions. No online support has existed to verify the correctness of drugs administered to the patients.

1.7 *Objectives of the research*

1. Create a Medical database to be used as an example to investigate negative associations between regular and frequent items while medical data moves in data streams.

2. Develop and implement an algorithm for mining medical data streams and a Pre-created Historical database and find the frequent and regular item sets and the negative associations existing in those item sets.

3. Find optimum frequencies and regularity thresholds that help find the most important negative associations.

2. Related work

Many techniques for mining frequent data sets from static databases are published. Data flows in streams with the Internet, especially from remote sensors to cloud storage. Dynamic mining follows data streams. To satisfy the user, respond quickly to constant requests. Lin et al. [13] utilized time-sensitive sliding windowing for frequent data stream mining. Their algorithm stores all frequent items in memory and a database containing expired data items. Based on storage space, the table can hold fewer expired data items.

One confidence threshold is typically used to identify positive and negative association rules ($A \Rightarrow B$, $A \Rightarrow \neg B$, $\neg A \Rightarrow B$, and $\neg A \Rightarrow \neg B$). In this method, the user must choose between positive and negative connections. Dong et al. [14] compared confidence criteria for four positive and three negative associations. The linkages between the four confidences showed that four confidence intervals should focus on the four types of associations. Considering the four confidences helps create deceptive rules. They explained how the chi-squared test mines association rules. They suggested a chi-squared test-based Positive and Negative Association Rules based on Multi-confidence (PNARMC) algorithm with four confidence thresholds.

Many data-gathering technologies have allowed every organization to store data on every activity. Large amounts of data make it hard to mine relevant information. Large data sets allow for more realistic and informative patterns. Sequential patterns give intriguing data insights. Existing sequential mining approaches focus on positive pattern behavior to anticipate the next event after a series. Anwar et al. [15] mine negative sequential patterns that contradict each other. Their technique finds patterned events/event sets.

Due to internet application migration, most data flows in streams. Streamed data is difficult to find fascinating. Support metric-based streamed data pattern mining is successful. However, pattern occurrence frequency is not a good measure for finding significant patterns. However, temporal regularity is best for online data stream mining for stock market applications. A pattern is regular if it appears within the user-defined period. None of the methods are frequently patterned from online application stream data. Regular Pattern Stream tree (RPS tree) and an effective mining method for detecting interesting patterns in streamed data were developed by Tanbeer et al. [16]. They recorded stream data in the RPS tree using a sliding window approach. The tree has been updated with the current data using an efficient approach.

Mirabedini et al. [17] reviewed all data stream pattern mining algorithms. Frequent item mining helps cluster and classify data. As network data volume has expanded, so needs to mine data streams for interesting patterns. Regular itemset mining is needed for static and streaming data. Frequent patterns might reveal business trends, scientific phenomena, etc. Pattern finding is the foundation for machine learning activities like association rule induction. Data has been scanned several times to uncover intriguing patterns. Only one scan of streamed data is possible; therefore, pattern-finding is performed in one scan. Since the stream length is unknown, such data has no closure. Before processing streaming data, an initial data set is collected and kept temporarily. VE: A review of online frequent pattern mining methods by Lee et al. [18]. They categorised approaches by pattern, data, and time window.

DSM-Miner by Yang et al. [19] mines maximal common patterns effectively. They have examined all data stream pattern mining techniques. Frequent item mining clusters and classifies data. Mining network data streams for intriguing patterns has increased as data volume has grown. Static and streaming data require consistent itemset mining. Frequent patterns may reflect business, scientific, etc., trends. Machine learning operations like association rule induction start with pattern detection. Data was scanned multiple times to find unusual patterns. Only one streamed data scan is possible; therefore, pattern-finding is done there. No closure exists for such data because the stream length is uncertain. A temporary data set is collected before processing streaming data.

Daly et al. [20] presented a structured review of online frequent pattern mining. By pattern, data, and period, they class approaches. Included the transaction sliding window mechanism that leverages each processing phase's transaction count. Decaying distinguishes old and new transactions. Sliding window maximum frequent pattern trees (SWM trees) are suggested for retaining frequent patterns. SWM tree root is used as an enumerated tree root for searching. Low interest or great confidence can make association rules stand out. The method assessed exceptional rule mining rules. They explored negative and exceptional association rules. Negative association rules generate exceptions. Also, they created a metric for anomalous rule interest. Candidate rules analyse patterns and decisions using exceptional rules and metrics.

Most literature-based tactics cut desirable decision-making patterns utilising fascinating criteria. However, choosing an interest measure is tricky and may need trial and error. No accurate method exists for determining interesting metrics. Thiruvady [21] suggested using user inputs to determine rules and limitations. The GRD algorithm finds the most intriguing rules.

Statistical correlations determine data set connectivity. Antonie and Zaiane [22] used a correlation between two item sets to identify negative association rules. Negative rules are retrieved if item sets correlate negatively, and confidence is strong.

Negative rules are retrieved if item sets are negatively correlated, and confidence is strong. Cornelis [23] examined many algorithms that mine negative and positive association rules and identified various failures. These characteristics were used to classify and catalogue mining algorithms and identify gaps. A confidence framework-based modified Apriori mining method can find negative correlations with intriguing ones. They used upward closure to match validity definitions' support-based negative connection interest. Dataset entries frequently have the intriguing "Support" parameter. Each level of data records has support values. Each level has multiple support values. The authors devised an Apriori-based upward closure method (PNAR) to find negative association rules.

Interesting negative and positive association rules (PNAR) and mining multiple-level support approaches have been developed. Their method mines positive and negative association rules from fascinating frequent and infrequent item sets using varying support levels. Based on item set regularity, Kumar and Rao [24] found positive and negative associations using vertical table mining.

Bagui and Dhar [25] demonstrated how to mine positive and negative association rules from MAP REDUCE data. They utilised the Apriori approach to mine several item sets, which was efficient but required much computation time.

None of the negative association rule mining studies Kisor and Porika [26], Ramasubbareddy et al. [27] used big data. Positive and negative association rules were found in uncommon item sets. The positive association rule mining identifies commonalities. This could reject many important or low-support items. Rare commodities or item sets can trigger negative association rules despite low support. Significant negative association rule mining requires more search space than positive rule mining for low-support objects. It would make sequential Apriori algorithm implementations easier and harder on big data. Few times has negative association rule mining been used.

Kaur et al. [28] reviewed AI-based illness diagnosis. However, no review has examined the medications' effects on heart disease prediction. Wei et al. [29] developed a machine learning-based drug response risk prediction algorithm. They anticipated disease-treatment drug risks. They have not considered patient risk when administering adversely related medications.

Lu et al. [30] suggested "Prefix-Span" sequential data mining and "Proportional Report Ratio" disproportionality-based technique to detect major adverse drug responses based on casual relationships, drugs, and drug reactions. They tested single drug-to-drug reactions. Restricting adverse pharmaceutical responses has not been studied.

Lu et al. [31] found predicted patterns in frequent item set mining. Negative correlations between suggested medicines are discovered. However, an uncommon item collection with negative drug connections is also important. The authors used bi-directional traversal to mine uncommon, closed item sets. However, negative relationships between infrequent or frequent item sets have not been studied.

Zhang et al. [32] developed a system to obtain drug interactions for patients with various conditions. They compiled a drug-drug interaction database from medical sources. Distance monitoring was used to extract drug-drug interactions. A transformer-based bidirectional encoder representation was employed to extract drug interactions. There is no modelling to classify positive or negative interactions.

Generalised tensor decomposition was used to create drug-gene-disease connections by Kim et al. [33]. Chemical structure and ATC code drug feature networks predicted drug-gene-disease relationships. They learned drug, gene, and sickness using a multi-layer perceptron neural network. They emphasised positive connections and disregarded unfavourable ones, especially among drugs.

Toor et al. [34] have presented a data stream mining method that includes enforcing Privacy before the same is disseminated. The data is mined from the data streams, and privacy is preserved. They primarily focussed on the drift in data Flowing in the streams due to changes in the data flow context.

Hewage et al. [35] have conducted a literature survey of the methods used for privacy-preserving data streaming mining methods. They focus primarily on the accuracy of privacy-preserving accuracy and data stream mining accuracy. They have not considered the issue of response time.

In his thesis, Mourtada [36] focused on developing a data stream association rule mining algorithm among co-occurring events, which means the kind of data stream varies occasionally. The proposed algorithm incrementally mines association rules over data streams that carry varied data.

Zhang et al. [37] have recommended a method to mine sequential patterns from data streams. They have used the sliding window method to arrive at the latest data by updating the old data with new data. A Prefix tree generates sequential patterns. They contended that negative associations can be mined using the Prefix tree.

Cuzzocrea et al. [38] have presented frequent itemset mining from dense data streams but have not concentrated on mining negative associations, which is the real problem in medical data. Lee et al. [39] have presented a "Path tree" algorithm for Mining sequential patterns efficiently in data stream environments. They have not focused on mining the negative associations. Tanbeer et al. [40] have presented a Sliding window-based frequent pattern mining over data streams but focused on mining positive patterns. This is a general-purpose presentation and does not focus on the medical domain or the data.

3. Establishing data streams

A medical database is created from registrations and prescriptions and fetching details from pharmaceutical companies regarding the chemical compositions of the drugs. 100,000 records have been used to create the database. This is synthetic data. Details such as Patient ID and Name have been masked to ensure the privacy of the data. The data is split into two parts (90,000 and 10,000) to serve as Historical and Stream Data.

One window is considered to have a data size equivalent to 1,000 records. Each window is 200 KB, with the record size being 200 Bytes. To get a new window, the window is continuously moved 1,000 Bytes to the left.

A buffer size of 10 Slots is considered, with each slot filled with a single data window. While one window is moved to store in the database, the adjacent slot is filled with a window of the data.

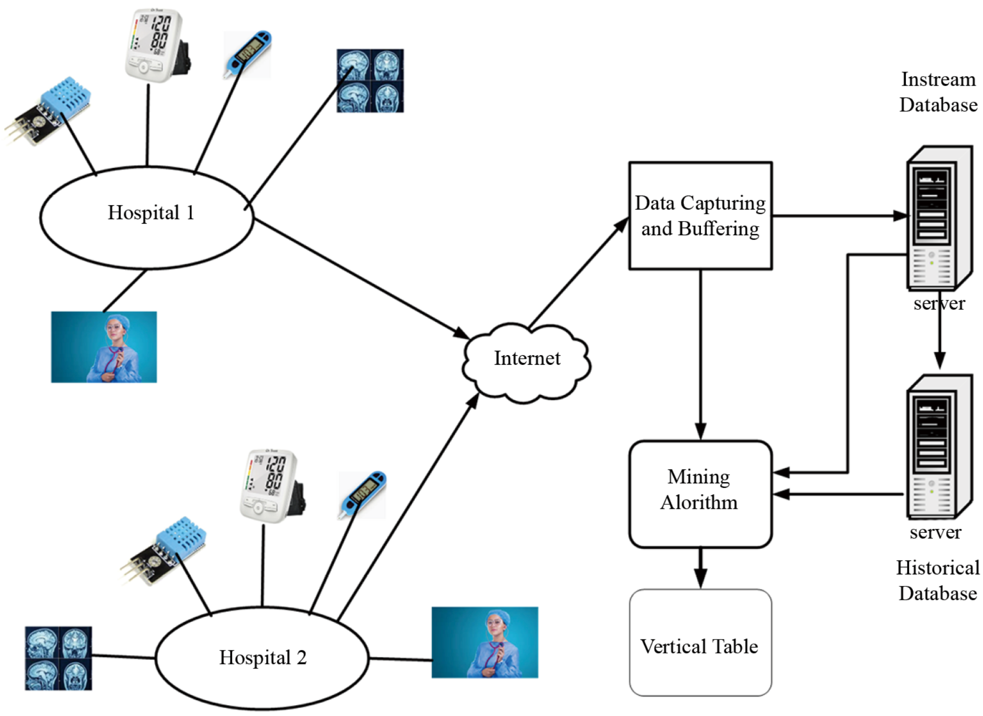

Figure 1. Capturing, buffering and creating a database representing the data streams

Figure 1 shows the data streaming method used to process ongoing and continuous data flow from medical establishments.

4.0 Streamed database algorithms for negative association finding based on item set regularity and frequency.

This technique generates negative patterns using sliding windowing. Data is scanned as received, and the initial data is copied into 10 windows, each window of size 1,000 records. 10 such windows are buffered, after which, while the latest window is received, the oldest window is written to a database.

A concurrent process accesses window database data. The Horizontal table, which lists item details and sales transactions, is updated. The pattern generation algorithm finds transaction data in instream data and writes it to a memory buffer. A buffer large enough to hold 1,000 transactions is allocated. After round-robin buffer scanning, the horizontal table was updated based on transaction table data.

From a historical database, Algorithm 1 creates an inverted table. Instream processing and Inverted Table updates are done in Algorithm 2. Algorithm 3 creates negative linkages between medications and their development chemicals.

Algorithm 1 Process-A

```
Read the support value to set pattern frequency and user-defined regularity thresholds.
Read historical DBMS data into memory, as indicated in Table 1.
Convert Table 2 data to vertical format; see Table 2.
```

Algorithm 2 Process-B

```
while TRUE do
    while there is an instream Transaction do
        if the Temp buffer is not full then
            Raise the pointer.
            Read the data and store it in the Temp Buffer using the pointer.
        else
            if the Window Buffer is full then
                Write the last window to the Stream Database, as shown in Table 3.
                for each of the records in the Last Window do
                    Find the frequent and regular item set.
                    if the item set satisfies the threshold values then
                        Add to inverted Table 4.
                        if the Item set is in the Inverted Table 2 then
                            Update the Transaction Table 5.
                            Update the Inverted table as shown in Table 6.
                        else
                            Add to the Inverted Table 6.
                        end if
                    end if
                end for
                Move the windows to the left.
                Write the temp-Buffer to the first window of the window buffer.
            else
                Move the Temp-Buffer to the first window of the stream Buffer.
            end if
        end if
    end while
    Prune the updated Inverted Table comparing with the threshold, implying deleting such records from Table 7.
end while
```

Algorithm 3 Process-C

while TRUE **do**
    # Find Negatively associated Chemicals Associations
    for every Next Record in **Table 8 do**
        Fetch the intersection of subsequent and current records if the intersection is empty.
        Find the frequency and regularity of the item sets.
        Enter the item sets into Table 8, if the frequency and regularity values satisfy the threshold values.
    end for
    # Find Negatively Associated Chemicals
    for each adversely correlated chemical in Table 8 **do**
        Discover related medications and indicate their negative correlations.
    end for
end while

Table 1. Sample medical data extracted from historical database (P-Patient, DE-Disease, DR = Drug, CH-Chemical in the drug)

P. SL. No.	Transaction ID	Patient number	Disease	Drug	Chemicals					Drug	Chemicals			
1	T2	P100	DE2	DR3	CH4	CH5	CH6	NA	NA	DR4	CH10	CH15	NA	NA
2	T4	P223	DE4	DR7	CH5	CH8	CH10	NA	NA	DR8	CH11	CH15	NA	NA
3	T6	P749	DE6	DR10	CH4	CH5	CH16	CH19	NA	NA	NA	NA	NA	NA
5	T8	P119	DE8	DR13	CH5	CH8	CH11	NA	NA	DR14	CH12	CH14	CH15	NA
	T10	P119	DE10	DR17	CH2	CH3	CH7	CH8	NA	DR18	CH13	CH14	CH15	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
	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
	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

Table 2. Inverted table along with filled-up frequency and regularity

Chemical code	Transaction ids																Maximum regularity (4)	Minimum frequency (3)
CH1		TT5		TT13	TT17	TT21											4	4
CH3			TT5				TT13		TT17	TT19	TT21						2	5
CH4		TT2	TT6	TT10	TT14	TT18											4	5
CH5		TT2	TT4		TT6	TT8		TT10	TT12	TT13	TT14	TT16	TT17	TT18	TT20	TT21	2	13
CH6	TT2		TT6	TT10	TT14		TT17	TT18									4	6
CH7	TT3																4	1
CH8	TT4	TT8			TT12	TT16	TT20										4	5
CH9		TT5		TT13	TT17	TT21											4	4
CH10		TT2	TT4	TT10	TT17												7	4
CH11	TT4		TT8	TT12	TT13	TT16	TT20										4	6
CH12		TT8		TT16	TT18	TT20											7	4
CH13	TT3																4	1
CH14			TT8		TT14	TT16	TT18	TT21									5	5
CH15	TT2		TT4	TT6	TT8	TT10	TT12	TT14	TT16	TT18	TT20						7	10

Table 3. Records from the data stream into the window buffer

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
	T3	P100	DE3	DR5	CH2	CH3	CH7	NA	NA	DR6	CH13	CH14	CH15	NA
	T5	P223	DE5	DR9	CH1	CH3	CH5	CH16	CH19	NA	NA	NA	NA	NA
4	T7	P937	DE7	DR11	CH2	CH3	CH7	CH11	NA	DR12	CH12	CH13	NA	NA
5	T9	P119	DE9	DR15	CH1	CH3	CH5	NA	NA	DR16	CH8	CH9	NA	NA
6	T11	P1235	DE11	DR19	CH5	CH8	CH11	CH15	NA	DR20	NA	NA	NA	NA
8	T15	P4573	DE15	DR27	CH1	CH3	CH5	NA	NA	DR28	CH9	CH11	NA	NA
10	T19	P10987	DE19	DR35	CH1	CH3	CH5	NA	NA	DR36	CH6	CH9	CH10	NA

Table 4. Inverted table for instream data

Chemical code	Transaction ids						
CH1	TT1	TT5	TT9	TT15	TT19		
CH2	TT1	TT3	TT7				
CH3	TT1	TT3	TT5	TT7	TT9	TT15	TT19
CH4	TT1						
CH5	TT1	TT5	TT9	TT11	TT15	TT19	

Table 4. (cont.)

Chemical code	Transaction ids			
CH6	TT19			
CH7	TT3	TT7		
CH8	TT6			
CH9	TT1	TT9	T15	T19
CH10	TT1	TT19		
CH11	TT11	TT15		
CH12	TT7			
CH13	TT3	TT7		
CH14	TT3			
CH15	TT3	TT11		
CH16	TT5			
CH19	TT5			

Table 5. Updated medical data (Historical data + instream data) (P-Patient, DE-Disease, DR = Drug, CH-Chemical in the drug)

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

Table 6. Updated inverted table along with filled-up frequency and regularity

Chemical code	Transaction ids	Maximum regularity (4)	Minimum frequency (3)
CH1	TT1 TT5 TT9 TT13 TT17 TT21	4	6
CH2	TT1 TT3 TT7 TT11 TT5 TT9	6	6
CH3	TT1 TT3 TT5 TT7 TT9 TT11 TT13 TT15 TT17 TT19 TT21	2	11
CH4	TT1 TT2 TT6 TT10 TT14 TT18 TT19	4	7
CH5	TT1 TT2 TT4 TT5 TT6 TT8 TT9 TT10 TT12 TT13 TT14 TT16 TT17 TT18 TT20 TT21	2	16
CH6	TT2 TT5 TT6 TT10 TT14 TT15 TT17 TT18	4	8
CH7	TT3 TT7 TT11 TT15 TT19	4	5
CH8	TT4 TT8 TT9 TT11 TT12 TT16 TT20	4	7
CH9	TT1 TT5 TT9 TT13 TT17 TT21	4	6
CH10	TT1 TT2 TT4 TT10 TT17	7	5
CH11	TT4 TT7 TT8 TT12 TT13 TT16 TT20	4	7
CH12	TT7 TT8 TT15 TT16 TT18 TT20	7	6
CH13	TT3 TT7 TT11 TT15 TT19	4	5
CH14	TT1 TT3 TT8 TT11 TT14 TT16 TT18 TT21	5	8
CH15	TT2 TT3 TT4 TT6 TT8 TT10 TT12 TT14 TT16 TT18 TT20	7	11

Table 7. Pruned inverted table for maximum regularity and minimum frequency

Chemical code	Transaction ids	Maximum regularity (4)	Minimum frequency (3)
CH1	TT1 TT5 TT9 TT13 TT17 TT21	4	6
CH3	TT1 TT3 TT5 TT7 TT9 TT11 TT13 TT15 TT17 TT19 TT21	2	11
CH4	TT1 TT2 TT6 TT10 TT14 TT18 TT19	4	7
CH5	TT1 TT2 TT4 TT5 TT6 TT8 TT9 TT10 TT12 TT13 TT14 TT16 TT17 TT18 TT20 TT21	2	16
CH6	TT2 TT5 TT6 TT10 TT14 TT15 TT17 TT18	4	8
CH7	TT3 TT7 TT11 TT15 TT19	4	5
CH8	TT4 TT8 TT9 TT11 TT12 TT16 TT20	4	7
CH9	TT1 TT5 TT9 TT13 TT17 TT21	4	6
CH11	TT4 TT7 TT8 TT12 TT13 TT16 TT20	4	7
CH13	TT3 TT7 TT11 TT15 TT19	4	5

Table 8. Drugs mapped to chemicals with negative association

Chemical	Associated drugs	Chemical	Associated drugs	Chemical	Associated drugs	Chemical	Associated drugs
CC4	DH3	CC8	DG7	CC11	DG11		
CC6	DG3	CC8	DG7	CC11	DG11		
CC4	DG3	CC6	DG3	CC8	DG7	CC11	DG11

4. Experimenting and results

4.1 *Creation of data set*

A database contains patent registration, diagnostic, patient-diagnosis, chemical, drug-chemical, quantity, and prescription codes. Hospitals have contributed 100,000 patient registrations and medications to the database. Each diagnosis, drug administration, and drug chemical makeup are included in an Example set. Data items in repeating groups are encoded and replaced with codes. We sort the sample set, compute the frequency and regularity of each item set, update the database, and import 100,000 Flat file entries. These records were processed using this paper's algorithm. No standard data set exists with the data items needed to discover negative connections.

5. Results

The procedure was applied on the data above, yielding the following results.

Step 1 Read the Historical Data Contained in a Database and add transaction IDs for each transaction. Details of the data fetched from the Database is shown in Table 1.

Step 2 Construct an Inverted Table in vertical format for the Transactions shown in Tables 1 and 2, along with the frequency of occurrence and the Regularity. Regularity is based on database record placements, and frequency is based on occurrence count.

Step 3 Read the Data from the Input data stream and load the same into a Temporary buffer. After filling the Temporary buffer, transfer the Records to window-1 of the window buffer after the last window is written to the database. The Records are read from the window Buffer. Table 3 shows the records from the data stream.

Step 4 Develop an Inverted Table from the buffered data. Table 4 shows the Inverted Table for records read from the window buffer. Here, the issue of frequency and regularity has not been used. The data is retrieved from online transactions.

Step 5 Update the Historical Inverted Table with the data in the inverted table relating to the data stream. The frequencies and the regularity are updated simultaneously. The updated Transaction and Inverted are shown in Tables 5 and 6.

Step 6 Prune the records which do not meet the Maximum Frequency and Minimum regulatory.

Records that do not satisfy the user's criterion are pruned using the maximum regularity and minimum frequency of 4 and 3, respectively. Table 3 reveals that Chemical codes CH2, CH10, CH12, CH14, and CH15 were trimmed because they do not fulfil the Regularity and Frequency criterion. Table 7 lists leftover records. By this criteria, 5 chemicals were pruned. Yellow records are trimmed because they don't meet thresholds.

Step 7 Make adverse chemicals. Find negative associations with records with no common transactions (Nill Common Items). Application of record intersection Produces unfavourable associations like (CC4 $\Rightarrow$ CC8, CC11), (CC4 $\Rightarrow$ CC8), (CC6 $\Rightarrow$ CC8), (CC4 $\Rightarrow$ CC11), (ACC6 $\Rightarrow$ CC11), (CC8 $\Rightarrow$ CC4, CC6), (CC6 $\Rightarrow$ CC8, CC11), (CC11 $\Rightarrow$ CC4, CC6), (CC4, CC6), $\Rightarrow$ CC8, CC11).

Step 8 Find Negatively Associated Drugs.

As stated in Table 8, map back the chemicals connected with the medications to determine the negatively associated pharmaceuticals.

The unfavourable associations show that DG3 should not be used with DG7 or DG11 because they conflict.

6. Discussion

Data analysis has been conducted considering different sample sizes of 20,000, 50,000 and 80,000 examples. The analysis is carried out considering thresholds in respect of frequencies and regularities. While the sample sizes are used for the historical database, the Instream data size is fixed at 10,000 examples. Table 9 shows the number of negative associations considering the sample size of 20,000 and data stream size of 10,000 examples.

Data gathering yielded 100,000 cases evaluated for 30,000 sample sizes 50,000, and 70,000. We set regularity and support thresholds and the number of unfavorable Associations observed. Table 9 shows frequent, regular, negative connections from 30,000 samples varying in regularity between 2.0 and 3.0 and frequency between 1.5 and 2.0.

Table 9. Negative frequent regular associations (20,000 historical and 10,000 data stream examples) vs. varying regularity + frequency

Total transactions	% Max regularity	% Support count	Number of negative frequent regular assertions
30,000	3	2	3
	3	1.75	21
	3	1.625	77
	3	1.5	230
30,000	2.5	1.75	20
	2.5	1.625	77
	2.5	1.5	176
	2.5	1.125	490
30,000	2	1.75	18
	2	1.625	59
	2	1.5	90
	2	1.1.25	334

Table 10 displays the number of negative Associations with frequency 1.75 and regularity 2-3 for 20,000 and 10,000 data stream instances. Negative associations respond to regularity modifications in Figure 1, with frequency fixed at 1.75. Figure 2 demonstrates the variation in negative correlations with regularity varying and Frequency set at 1.75.

Table 10. Negative associations vs. frequency @ 1.75 and varying regularity (20,000 historical + 10,000 data stream examples)

Regularity	Negative frequent regular item sets	% of negative associations * 1,000
3.00	3	3/30,000 = 0.10
2.50	20	20/30,000 = 0.66
2.00	18	18/30,000 = 0.60
The average percentage of negative association per 1,000 examples		0.45

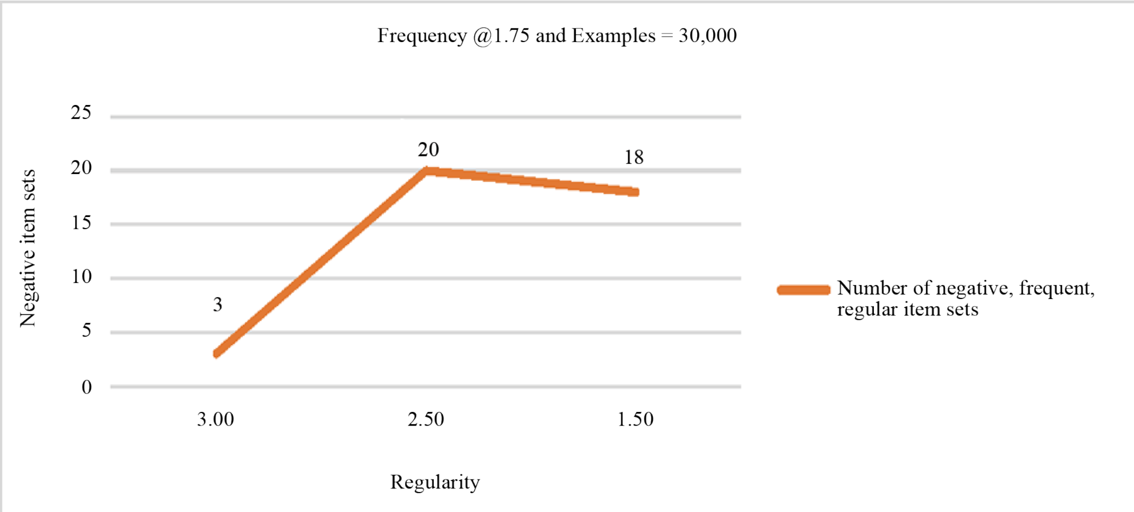

Figure 2. Negative association item sets @ frequency (1.75) and (30,000 examples) and varying regularity

Table 11 illustrates the number of negative connections with frequency 1.5 and regularity 2-3 in 30,000 cases. The line graph of negative associations with 1.5 frequency and variable regularity is shown in Figure 2. Figure 3 shows the variation in negative associations with regularity varying and frequency set at 1.50.

Table 11. Negative associations @ frequency at 1 (30,000 examples) varying regularity

Regularity	Negative frequent regular item sets	% of negative associations * 1,000
3.00	230	230/30,000 = 7.6
2.50	176	176/30,000 = 5.8
2.00	90	90/30,000 = 3.0
Average negative associations per 1,000 examples		5.4

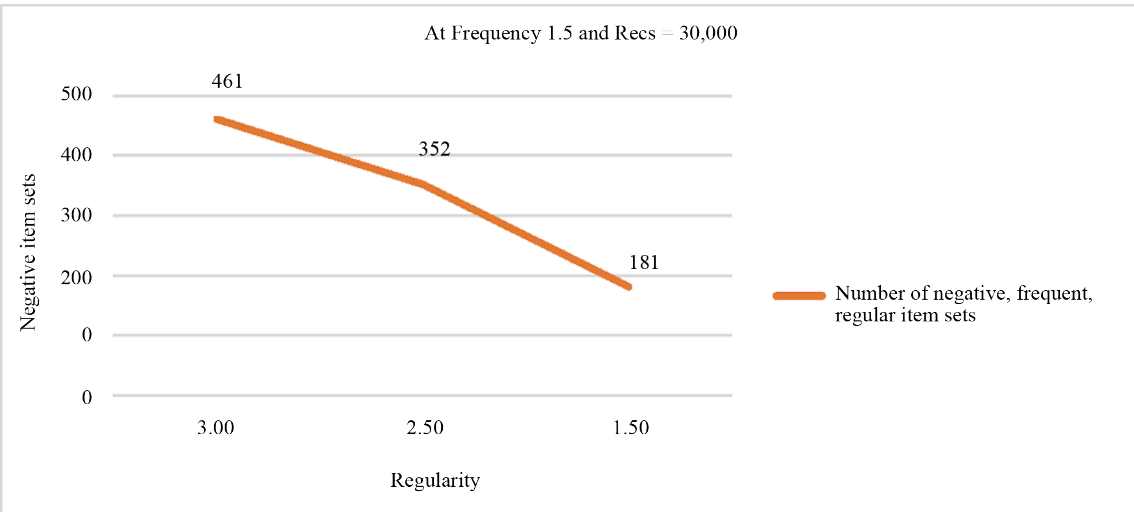

Figure 3. Negative association @ regularity (1.50) (30,000 examples) varying the frequency

Figure 3 negative association item sets @ frequency (1.50) and (30,000 examples) varying regularity.

The larger example set was analysed further. Table 12 illustrates the percentage of negative correlations per 1,000 examples for sample sizes (20,000, 50,000, 80,000), including 10,000 DataStream instances. Frequency varies.

Table 12. % reduction in negative relationships with larger samples and appropriate frequency and support

Total transactions	% Max regularity	% Support count	Number of negative frequent regular	% of negative associations * 1,000	Average % of negative associations per 1,000 samples
30,000	1.50	1.750	18	18/30 = 0.60	1.85
	1.50	1.625	59	59/30 = 1.96	
	1.50	1.500	90	90/30 = 3.0	
50,000	1.65	1.65	7	7/50 = 0.14	0.60
	1.65	1.25	20	20/50 = 0.40	
	1.65	1.00	75	75/50 = 1.25	
80,000	1.35	1.65	18	18/80 = 0.22	0.70
	1.35	1.35	59	59/80 = 0.73	
	1.35	1.00	90	90/80 = 1.13	

From the above Table, it can be inferred that Considering a sample size of 50,000 examples, fixing the regularity 1.65 and frequency 1.25 gives the least and most effective negative associations 20 out of 50,000 could be identified.

Further analysis examines the effects of fixing regularity, altering frequency, and different sample sizes. Table 13 lists unfavourable relationships with regularity = 1.50 and sample size = 30,000. Figure 4 exhibits negative association variation.

Table 13. Negative associations @ regularity 1.50 (30,000 examples) varying frequency

Frequency	Negative frequent regular item sets	% of negative associations * 1,000
1.750	18	18/30 = 0.60
1.625	59	59/30 = 1.97
1.500	90	90/30 = 3.00
	Average	1.85

Table 14 lists negative relationships with regularity 1.65 and sample size 50,000. Negative connections dropped 73%. Figure 5 shows variances.

Table 14. Negative associations for a sample size of 50,000 fixing the regularity @ 1.65

Frequency	Negative frequent regular item sets	% of negative associations * 1,000
1.650	7	7/50 = 0.14
1.250	20	20/50 = 0.40
1.000	75	75/50 = 1.5
	Average	0.68

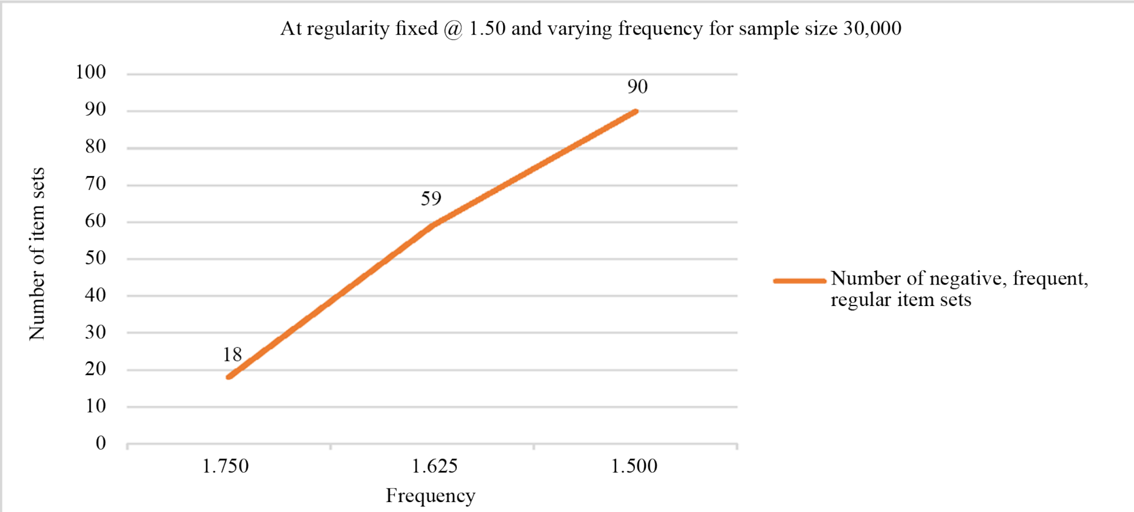

Figure 4. Negative association item sets @ regularity (1.65) (examples 50,000) varying the frequency

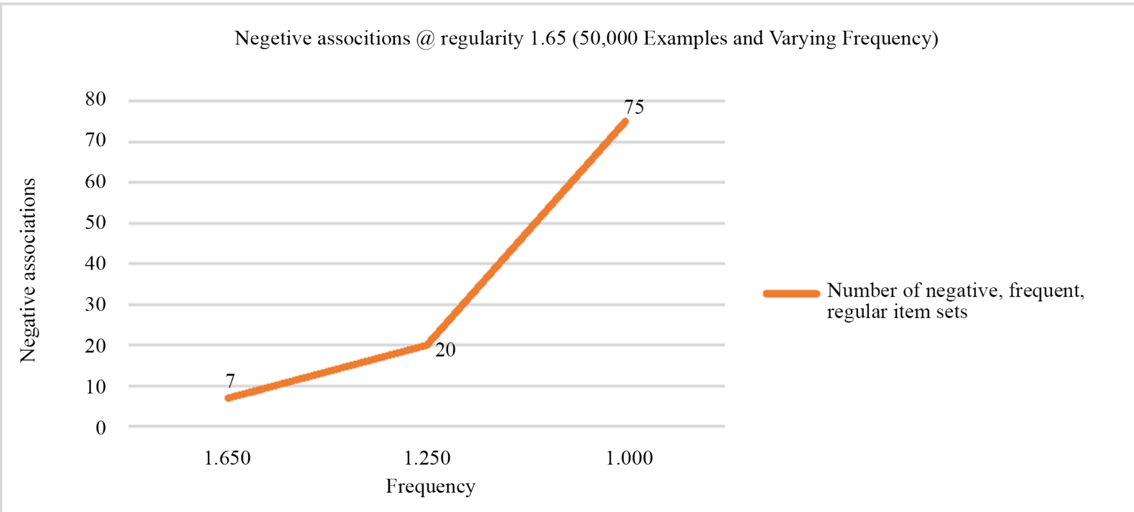

Figure 5. Negative association item sets @ regularity (1.65) (examples 50,000) varying the frequency

In both cases (30,000 examples and 50,000 examples), when regularity is fixed, and frequency is varied, the number of Negative associations is the least when the highest frequency is chosen. The negative associations of 0.68 per 1,000 samples can be achieved when the sample size is 50,000, with regularity fixed at 1.65.

6.1 *Comparative analysis*

Two algorithms Minimum weighted Association rules-Sliding windows (MWAR-SW) and Minimum positive and negative association rules (MPNAR-SW)-Sliding Window are related to mining association rules from Transactional data set T154D1000K which is a E-commerce data.

These algorithms only find the direct associations either positive or negative or both. They don't mine the derived negative associations like chemical compositions presented in this paper. To derive comparisons, we have run these algorithms against medical database and comparison made as shown in Table 15. From the table it can be seen only algorithm presented in this paper is focussed at negative and derived negative associations. This paper is first of its kinds that runs on synthetic medical database. The algorithm presented in this paper produces optimum results at low frequency (1.75) while other papers have not considered the issue of regularity. The time taken to process 30,000 transaction is minimum (175 s). When run on i9, 8core intel, and 2GPU NIVIDIA.

Table 15. Comparative analysis

Parameter	MWAR-SW [42]	MPNAR-SW [41]	This paper
Are the positive associations considered?	Yes	Yes	No
Are the negative associations considered?	No	Yes	Yes
Are the derived negative associations found?	No	No	Yes
Is regularity uses as interesting measure?	No	No	Yes
Is the frequency used as an interesting measure?	Yes	Yes	Yes
Performance in microseconds for processing 30,000 transactions.	400 s	337	175 s
Maximum frequency at which optimum results have been obtained.	2.35	2.30	1.75
Maximum regularity at which optimum results have been obtained.	Not applicable	Not applicable	1.50
Number of negative associations found.	0	234	600

7. Conclusions and future scope

Medical data moves in streams, and doctors must be warned soon after prescription if any negative associations exist among drugs prescribed by them, or else there could be a devastating effect on the patients. Many patients died during the coronavirus time as no system existed at that time that warned doctors not to use some drugs in combination. The use of humidifier drugs killed many due to many side reactions that affected the human organs.

Finding minimum and effective negative correlations is critical when medical data is rapidly shared online. Critical items are emphasised. Minifying regular and common item sets from historical databases and data streams yields the most effective negative associations. Since the complete data set is unavailable for study, finding the maximal and closed data sets is impossible. Regularity at 1.5 and frequency at 1.75 yields the last and most effective negative associations.

This research identifies all negative correlations with a collection of medications given to patients. Finding positive associations can be challenging due to many negative ones. Every medicine is chemically made. Chemical mixtures can be harmful. Consider negative correlations among chemicals and decode them into medications to determine the most important ones. This research is mainly limited by the chemical composition of medications doctors give to treat disorders.

Research medication administration limits and large trends. Rare and new medicinal item sets must be investigated. Further research may reveal intriguing elements beyond frequency and support that directly fit to uncover effective negative associations. The method may find negative links in remote or incremental medical databases and streaming data.

Conflict of interest

The authors declare no competing financial interest.

References

[1] Ramzan F, Ayyaz M. A comprehensive review on data stream mining techniques for data classification and future trends. *EPH-International Journal of Science And Engineering*. 2023; 9(3): 1-29. Available from: https://doi.org/10.53555/ephijse.v9i3.201.

[2] ManaL M, Manal A. Mining techniques for streaming data. *International Journal of Data Mining & Knowledge Management Process*. 2022; 12(2): 1-14. Available from: https://doi.org/10.5121/ijdkp.2022.12201.

[3] Burattin A. Streaming process mining. In: van der Aalst WMP, Carmona J. (eds.) *Process Mining Handbook*. Cham, Switzerland: Springer; 2022. p.1-29. Available from: https://doi.org/10.1007/978-3-031-08848-3_11.

[4] Mahmood S, Shahbaz M, Guergachi A. Negative and positive association rules mining from text using frequent and infrequent item sets. *Science World Journal*. 2014; 2014: 973750. Available from: https://doi.org/10.1155/2014/973750.

[5] Aggarwal CC, Yu PS. Mining associations with the collective strength approach. *IEEE Transactions on Knowledge and Data Engineering*. 2001; 13(6): 863-873. Available from: https://doi.org/10.1109/69.971193.

[6] Aggarwal CC, Yu PS. A new framework for item-set generation. In: *Proceedings of the Seventeenth ACM SIGACT-SIGMOD-SIGART Symposium on Principles of Database Systems (PODS'98)*. NY, USA: ACM Press; 1998. p.18-24. Available from: https://doi.org/10.1145/275487.275490.

[7] Agrawal R, Imielinski T, Swami A. Mining association rules between sets of items in large databases. In: *ACM SIGMOD Record*. New York, NY, USA: ACM Press; 1993. p.207-216. Available from: https://doi.org/10.1145/170036.170072.

[8] Agrawal R, Srikant R. Fast algorithms for mining association rules. In: *Proceedings of the 20th International Conference on Very Large Data Bases (VLDB'94)*. San Francisco, CA, USA; 1994. p.487-499.

[9] 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: IEEE; 1998. p.494-502. Available from: https://doi.org/10.1109/ICDE.1998.655812.

[10] Zaki MJ. Fast vertical mining using diffsets. In: *Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'03)*. New York, NY, USA: ACM Press; 2003. p.326-335. Available from: https://doi.org/10.1145/956750.956788.

[11] Yang CY, Li Y, Zhang C, Hu Y. A novel algorithm of mining maximal frequent pattern based on projection sum tree. In: *Proceedings of the Fourth International Conference on Fuzzy Systems and Knowledge Discovery (FSKD 2007)*. Haikou, China: IEEE; 2007. p.458-462. Available from: https://doi.org/10.1109/FSKD.2007.95.

[12] Gulzar K, Ayoob Memon M, Mohsin SM, Aslam S, Akber SMA, Nadeem MA. An efficient healthcare data mining approach using apriori algorithm: a case study of eye disorders in young adults. *Information*. 2023; 14(4): 203. Available from: https://doi.org/10.3390/info14040203.

[13] Lin CH, Chiu DY, Wu YH, Chen ALP. Mining frequent itemsets from data streams with a time-sensitive sliding window. In: *Proceedings of the Society for Industrial and Applied Mathematics (SIAM) Conference*. Newport Beach, CA, USA; 2006. p.1-10.

[14] Dong X, Sun F, Han X, Hou R. Study of positive and negative association rules based on multiconfidence and chi-squared test. In: *Advanced Data Mining and Applications (ADMA'06)*. Berlin, Germany: Springer-Verlag; 2006. p.100-109. Available from: https://doi.org/10.1007/11811305_10.

[15] Anwar F, Petrounias I, Morris T, Kodogiannis V. Discovery of events with negative behavior against given sequential patterns. In: *Proceedings of the 2010 5th IEEE International Conference Intelligent Systems*. London, UK: IEEE; 2010. p.1-8. Available from: https://doi.org/10.1109/ICDMW.2010.14.

[16] Tanbeer SK, Ahmed CF, Jeong BS. Mining regular patterns in data streams. In: *Database Systems for Advanced Applications (DASFAA'10)*. Berlin, Germany: Springer-Verlag; 2010. p.399-413. Available from: https://doi.org/10.1007/978-3-642-12026-8_31.

[17] Mirabedini S, Panah MA, Darbanian M. Frequent pattern mining in data streams. *Journal of Engineering and Applied Sciences*. 2017; 12(4): 857-863.

[18] Lee VE, Jin R, Agarwal G. Frequent pattern mining in data streams. In: *Frequent Pattern Mining*. Cham, Switzerland: Springer International Publishing; 2014. p.1-20. Available from: https://doi.org/10.1007/978-3-319-07821-2_9.

[19] Yang J, Wei Y, Zhou F. An efficient algorithm for mining maximal frequent patterns over data streams. In: *2015 7th International Conference on Intelligent Human-Machine Systems and Cybernetics*. Hangzhou, China: IEEE; 2015. p.444-447. Available from: https://doi.org/10.1109/IHMSC.2015.226.

[20] Daly O, Taniar D. Exception rules mining based on negative association rules. In: *Computational Science and Its Applications-ICCSA 2004*. Berlin, Germany: Springer-Verlag; 2004. p.543-552. Available from: https://doi.org/10.1007/978-3-540-24768-5_58.

[21] Thiruvady DR, Webb GI. Mining negative association rules using GRD. In: *Proceedings of the Pacific-Asia Conference on Knowledge Discovery and Data Mining*. Sydney, Australia; 2004. p.161-165. Available from: https://doi.org/10.1007/978-3-540-24775-3_20.

[22] Antonie M, Zaiane OR. 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; 2004. p.27-38. Available from: https://doi.org/10.1007/978-3-540-30116-5_6.

[23] 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: IEEE; 2006. p.1-6. Available from: https://doi.org/10.1109/ICCIS.2006.252251.

[24] Kumar NVSP, Rao KR. Mining positive and negative regular item-sets using vertical databases. *International Journal of Simulation Systems, Science and Technology*. 2016; 17: 3. Available from: https://doi.org/10.5013/IJSSST.a.17.32.33.

[25] Bagui S, Dhar PC. Positive and negative association rule mining in Hadoop's MapReduce environment. *Journal of Big Data*. 2019; 6: 75. Available from: https://doi.org/10.1186/s40537-019-0238-8.

[26] 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)*. Coimbatore, India: IEEE; 2016. p.1-5. Available from: https://doi.org/10.1109/INVENTIVE.2016.7823240.

[27] Ramasubbareddy B, Govardhan A, Ramamohanreddy A. Mining positive and negative association rules. In: *2010 5th International Conference on Computer Science & Education*. Hefei, China: IEEE; 2010. p.1403-1406. Available from: https://doi.org/10.1109/ICCSE.2010.5593755.

[28] Kaur S, Singla J, Nkenyereye L, Jha S, Prashar D, Joshi GP, et al. Medical diagnostic systems using artificial intelligence (AI) algorithms: Principles and perspectives. *IEEE Access*. 2020; 8: 228049-228069. Available from: https://doi.org/10.1109/ACCESS.2020.3042273.

[29] Wei J, Lu Z, Qiu K, Li P, Sun AH. Predicting drug risk level from adverse drug reactions using SMOTE and machine learning approaches. *IEEE Access*. 2020; 8: 185761-185775. Available from: https://doi.org/10.1109/ACCESS.2020.3029446.

[30] Lu Y, Chen S, Zhang H. Detecting potential serious adverse drug reactions using sequential pattern mining method. In: *2018 IEEE 9th International Conference on Software Engineering and Service Science (ICSESS)*. Beijing, China: IEEE; 2018. p.56-59. Available from: https://doi.org/10.1109/ICSESS.2018.8663856.

[31] Lu Y, Seidl T. Towards efficient, closed infrequent item set mining using bi-directional traversing. In: *2018 IEEE 5th International Conference on Data Science and Advanced Analytics (DSAA)*. Turin, Italy: IEEE; 2018. p.140-149. Available from: https://doi.org/10.1109/DSAA.2018.00024.

[32] Zhang J, Liu W, Wang P. Drug-drug interaction extraction from Chinese biomedical literature using distant supervision. In: *2020 IEEE International Conference on Knowledge Graph (ICKG)*. Nanjing, China: IEEE; 2020. p.593-598. Available from: https://doi.org/10.1109/ICBK50248.2020.00089.

[33] Kim Y, Cho Y. Predicting drug-gene-disease associations by tensor decomposition for network-based computational drug repositioning. *Biomedicines*. 2023; 11: 1998. Available from: https://doi.org/10.3390/biomedicines11071998.

[34] Toor AA, Usman M, Younas F, Fong ACM, Khan SA, Fong S. Mining massive e-health data streams for IoMT enabled healthcare systems. *Sensors*. 2020; 20: 2131. Available from: https://doi.org/10.3390/s20072131.

[35] Hewage UHW, Sinha R, Naeem MA. Privacy-preserving data (stream) mining techniques and their impact on data mining accuracy: A systematic literature review. *Artificial Intelligence Review*. 2023; 56: 10427-10464. Available from: https://doi.org/10.1007/s10462-023-10425-3.

[36] Mourtada AF. *Mining Association Rules Events Over Data Streams*. Master Thesis. Concordia Institute for Information Systems Engineering, Concordia University; 2017.

[37] Zhang N, Ren X, Dong X. An effective method for mining negative sequential patterns from data streams. *IEEE Access*. 2023; 11: 31842-31854. Available from: https://doi.org/10.1109/ACCESS.2023.3262823.

[38] Cuzzocrea A, Jiang F, Lee W, Leung CK. Efficient frequent itemset mining from dense data streams. In: *Proceedings of the Asia-Pacific Web Conference*. Changsha, China; 2014. p.593-601. Available from: https://doi.org/10.1007/978-3-319-11116-2_56.

[39] Lee G, Hung K-C, Chen Y-C. Path tree: Mining sequential patterns efficiently in data streams environments. In: *Proceedings of the Advances in Intelligent Systems and Applications*. Taiwan; 2013. p.261-268. Available from: https://doi.org/10.1007/978-3-642-35452-6_28.

[40] Tanbeer SK, Ahmed CF, Jeong B-S, Lee Y-K. Sliding window-based frequent pattern mining over data streams. *Information Sciences*. 2009; 179(22): 3843-3865. Available from: https://doi.org/10.1016/j.ins.2009.07.012.

[41] Ouyang W. Mining positive and negative association rules in data streams with a sliding window. In: *2013 Fourth Global Congress on Intelligent Systems*. Hong Kong, China: IEEE; 2013. p.205-209. Available from: https://doi.org/10.1109/GCIS.2013.39.

[42] Ouyang W, Huang Q. Mining weighted association rules in data streams with a sliding window. In: *2015 International Conference on Computer Science and Intelligent Communication*. Atlantis Press; 2015. p.271-274. Available from: https://doi.org/10.2991/csic-15.2015.65.