



# Rare pattern mining: challenges and future perspectives

Anindita Borah<sup>1</sup> · Bhabesh Nath<sup>1</sup>

Received: 3 June 2018 / Accepted: 19 October 2018 / Published online: 10 November 2018  
© The Author(s) 2018

## Abstract

Extracting frequent patterns from databases has always been an imperative task for the data mining community. Literature has endowed plentiful endeavors to this research area with significant breakthroughs. Mining of rare patterns although being subsided has proved to be of vital importance in many domains. The application of rare pattern mining is inevitable and thus has become an emerging field of research. However, discovering rare patterns from databases comes up with numerous challenges. This article provides a concise overview of the various research issues involved in rare pattern mining through experimental analysis using real-life and synthetic datasets. Rare pattern mining being a new area has abundant scope and there are still certain gaps that need to be filled. In this article, we also present some viable future directions for the researchers to better perceive the area of rare pattern mining.

**Keywords** Rare patterns · Frequent patterns · Rare item dilemma · Challenges

## Introduction

Rare patterns, unlike the frequent ones, are those whose frequency of appearance in the dataset is below a user-defined threshold. Frequent pattern mining techniques tend to prune such patterns considering them to be undesirable or of no interest. The research community, however, has witnessed the significance of rare patterns in many domains. For instance, inimical drug reactions can be identified by some rare responses to medications in the field of biology. Similarly in the field of network security, rare events or occurrences may indicate some security threats or network failures. Mining rare patterns using traditional frequent pattern mining techniques proves to be ineffectual if the user-defined threshold is pushed too low, an issue known as rare item dilemma. Existing frequent pattern generation methods for rare pattern mining may spawn enormous number of patterns or rules escalating the computational complexity. Thus significant rare pattern mining techniques have been devised for extracting the rare patterns.

Many significant works have been reported in the area of rare pattern mining in recent years. The different endeavors for mining rare patterns have extensively employed the eminent pattern mining strategies like Apriori and FP-Growth. Since its inception, there have been a wide range of research publications addressing the various issues involved in the extraction of rare patterns. Despite such numerous and fruitful attempts, there are still some issues that demand utmost attention from the rare pattern mining community. This flourishing field thus appeals for an exhaustive review of the various issues and challenges associated with the mining of rare patterns and some feasible solutions for eradicating the same as future directions for the researches. Although there is an initial attempt to provide the literature review of existing rare pattern mining techniques in [77], till now no initiative has been taken to outline the major rare pattern mining challenges through experimental analysis along with significant future perspectives for the same.

In this article, we attempt to contribute an extensive review on the obstacles encountered during rare pattern mining through experimental analysis using benchmark datasets, along with possible solutions. There is immense scope for this emerging area and some significant issues in the field of pattern mining is still untouched. This paper thus aims to provide some future directions worthwhile for the researchers. To sum up, the major contributions of this paper are as follows:

✉ Anindita Borah  
anindita01.borah@gmail.com

Bhabesh Nath  
bnath@tezu.ernet.in

<sup>1</sup> Department of Computer Science and Engineering, Tezpur University, Napaam, Tezpur, Sonitpur, Assam 784028, India

- Identification of the major rare pattern mining challenges through experimental analysis using real-life and synthetic datasets.
- Comparison between the area of frequent and rare pattern mining with respect to the number of initiatives taken.
- Illustration of significant future directions for the area of rare pattern mining.

Our discussion on this paper is arranged as follows: Significance of rare patterns and rare rules is depicted in “Significance of rare patterns and rare association rules”. “Rare pattern mining methodologies” provides a brief illustration of different methodologies for mining rare patterns followed by experimental analysis of the major research challenges faced by rare pattern mining techniques in “Major research challenges for rare pattern mining techniques”. A comparison between the number of attempts based on frequent and rare pattern mining is elucidated in “Frequent vs rare pattern mining: a comparison”. The paper proceeds to discuss some viable future directions for the rare pattern mining community in “Future directions for rare pattern mining” and finally ends with a conclusion in “Conclusion”.

## Significance of rare patterns and rare association rules

Mining of rare patterns from databases has always been overlooked, giving more emphasis on the frequent ones. Recent studies show that these uncommon or unusual patterns are proficient in discovering hidden useful information from databases in various domains. Significance of rare patterns and the rare association rules obtained from them is manifold. There are some areas where the rare patterns have been found to be more important as compared to the frequent ones. This section attempts to establish the importance of rare patterns and rare association rules in different domains with the help of suitable examples. Following are some of the application areas where detection of rare patterns may prove to be beneficial over the frequent patterns:

1. *Network intrusion detection* In case of network packet databases for Intrusion Detection Systems (IDS), the number of intrusions is very less as compared to the total network traffic. The intrusions thus represent some rare events that need to be considered for detecting network anomalies.
2. *Credit card fraud detection* The credit card transactions stored in databases are millions in number. However, fraudulent activities in such transactions are rare and hence are very few in number.
3. *Medical diagnosis* In medical diagnosis, mammogram images are often used for cancer detection. In the entire

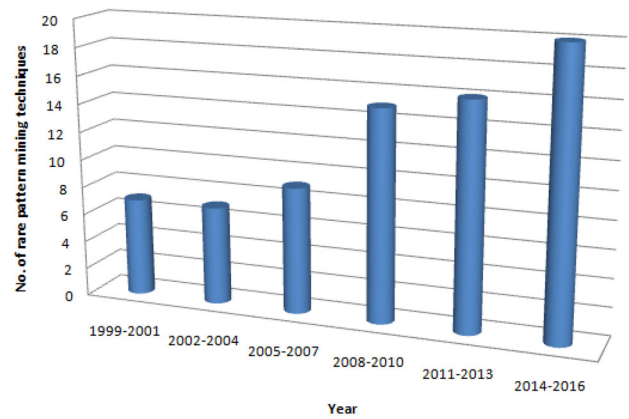


Fig. 1 Number of rare pattern mining techniques developed

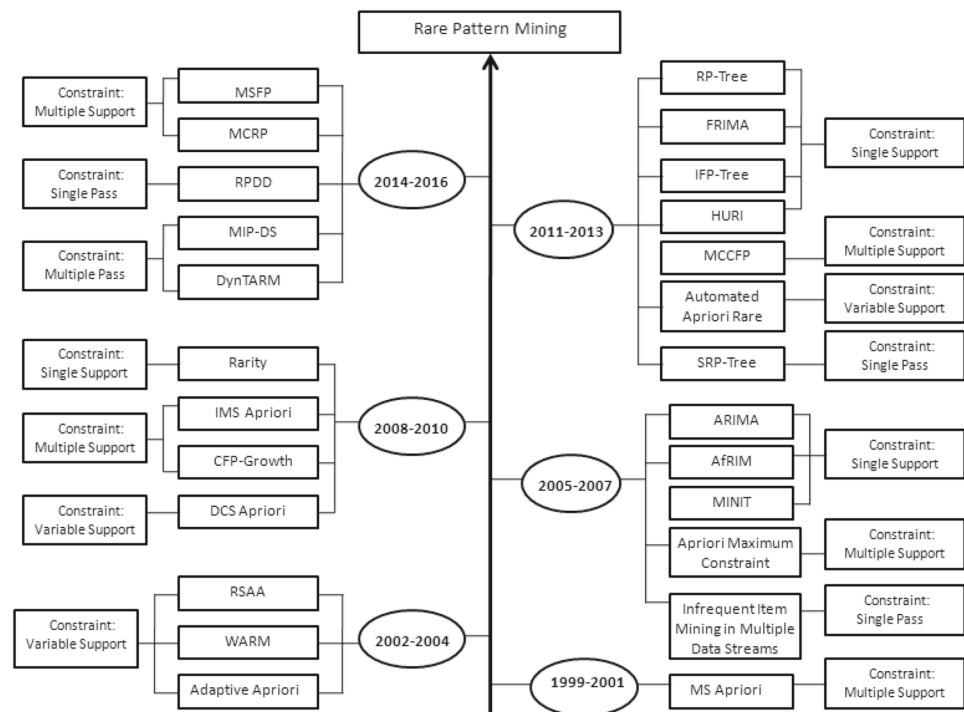
image, however, only a small fragment indicates the cancerous pixels.

4. *Insurance risk modeling* In case of insurance companies, the claims from the insurers are rare but may prove to be costly for them.
5. *Web mining* In on-line marketing applications, although a lot of people visit the website, only a small percentage of the people make the purchase.
6. *Hardware fault detection* Faults in hardware equipments occur very rarely but need to be considered for the detection of equipment failure.

The research community is greatly benefited by the emergence of rare pattern mining. Over the years, there has been considerable growth and progression of the area of rare pattern mining. The fact becomes evident from Fig. 1 that justifies the growing interest of researchers for rare patterns and rare rules extracted from them. The figure is a graphical illustration of the number of rare pattern mining techniques developed since 1999, spanning different issues of rare pattern extraction.

## Rare pattern mining methodologies

The concept of pattern mining was first introduced by Agrawal et al. [8], for mining the frequent patterns. They defined  $K = \{k_1, k_2, \dots, k_m\}$  as a set of items and insisted that the itemset  $K$  is frequent if and only if its frequency of occurrence in the database  $D$  is equal to or greater than the user-defined minimum support threshold. Their initial endeavor towards frequent pattern mining is the eminent Apriori algorithm that employs a downward closure property for producing the frequent patterns. The Apriori property illustrates that “an itemset  $K$  is frequent only if all its subsets are frequent”. Based on this property, the database is scanned to generate the frequent 1-itemsets which are further used to

**Fig. 2** Evolution of rare pattern mining techniques

generate frequent 2-itemsets and so on. This process continues until no more itemsets can be generated.

The efficiency of Apriori algorithm is considerably reduced by the generation of enormous candidate itemsets and multiple scanning of the database. To overcome this, Han et al. [50] developed a data structure-based method for mining the frequent patterns in mere two database scans. A tree data structure called FP-Tree was used to maintain the information of the database that prevents scanning the database time and again. There after only the FP-Tree can be employed during the mining process instead of referring the database every time. This greatly minimizes the computational complexity by decreasing the number of database scans and producing the frequent patterns without generating candidates. The Apriori and FP-Growth algorithms were further extended for allowing the retainment of rare patterns. Many of their variants for mining rare patterns have been introduced in the literature.

The domain of rare pattern mining have greatly evolved since its inception, spawning many Apriori and FP-Tree based approaches. Figure 2 provides a year-wise depiction of different approaches developed for rare pattern mining along with the constraints adopted. This section elucidates some attempts for mining the rare patterns using the frequent pattern mining methodologies. “Extensions of apriori” illustrates some versions of Apriori for rare pattern mining while “Extensions of FP-growth” discusses the FP-Growth extensions.

## Extensions of apriori

As discussed earlier, the primer algorithm in the field of pattern mining is the Apriori algorithm. Apriori algorithm is suitable for frequent pattern mining and cannot be used precisely for mining the rare patterns. Hence many variations of Apriori have been proposed for rare pattern mining.

The first attempt towards rare association rule mining was made by Liu et al. [96] in their algorithm called MS-Apriori that employs an Apriori like strategy to incorporate some rare items during itemset generation. The authors argued that a single support threshold cannot be used for extracting the rare patterns effectively and ended up proposing a “multiple support framework” for the same. The framework assigns each item their individual support values instead of relying on a single one. The algorithm is efficient in finding rare patterns but it employed an additional parameter  $\beta$  that adds to the computational complexity of the algorithm. Kiran et al. [74] in their algorithm IMS-Apriori, improved the initial MS-Apriori algorithm by incorporating another parameter of support difference. Even though it succeeded in generating more number of rare items it increases the burden of assigning two extra parameters:  $\beta$  and support difference. Lee et al. [85] extended the concept of multiple minimum supports using a model called maximum constraints model. The minimum support considered in this case is the maximum value among the minimum support values assigned to each item. The algorithm is faster due to granular bit string computation but fails to generate the complete set of rare items.

Some algorithms extend the Apriori algorithm and use only a single minimum support threshold to find the rare itemsets. The most significant effort in this regard was made in [135]. The algorithm called ARIMA is capable of finding the complete set of rare items but spends a lot of time looking for the rare and frequent itemsets. ARIMA is further extended by Hoque et al. [56] in their algorithm FRIMA that generates both the frequent and rare itemsets. The algorithm maintains the rare, frequent and zero itemsets in three different candidate lists and later on merges the lists containing frequent and rare itemsets into a single list removing the zero itemsets. The algorithm managed to generate the complete set of rare items in lesser execution time than ARIMA but consumed a higher amount of memory due to the retainment of zero itemsets along with the frequent and rare itemsets. Adda et al. [2] employed a strategy different from the previous approaches. Their algorithm AFRIM, performs the level-wise search in top-down fashion unlike the traditional bottom-up search approach. The algorithm initially generates the largest candidate itemset combining all rare items and then proceeds to generate the smaller candidate itemsets. Similar to FRIMA, it also suffers from the drawback of generating zero support itemsets. Pillay and Vyas [120] identified the need for high-utility rare itemsets and proceed to generate the same in their algorithm HURI. To measure the significance of rare itemsets, HURI consider the utility values of the itemsets along with their frequencies. The itemsets satisfying the predefined minimum utility value are considered to be rare, discarding other itemsets. Despite generating the user interested rare itemsets, the algorithm proves to be tedious due to the pre-assignment of utility values to each individual item. Instead of generating the complete set of rare itemsets, Haglin and Manning [45] developed the MINIT algorithm to generate only a subset of the rare itemsets called minimal infrequent itemsets. The algorithm assigns individual ranks to the items based on their support values and further considers only the higher rank items for itemset generation. The algorithm spends lesser execution time due to the generation of only minimal infrequent itemsets but still misses out some significant rare itemsets. Rarity algorithm proposed by Troiano et al. [144] considers the longest transaction in the database for rare itemset generation and performs a level-wise top-down search like AFRIM. The algorithm maintains a Candidate list for retaining the rare itemsets and a Veto list for retaining the frequent itemsets. The rare itemsets generated are finally stored in another list. Despite generating the complete set of rare items, the algorithm undergoes memory overhead.

In addition to the usage of a single minimum support threshold or multiple minimum support thresholds, some rare pattern mining techniques employ dynamic thresholds or more than one threshold. RSAA algorithm proposed by Yun et al. [166] employed two thresholds, one for generat-

ing the rare itemsets and another for generating the frequent itemsets. The advantage of this algorithm is that it is independent of the parameter  $\beta$  employed by MS-Apriori but fails to outperform in terms of execution time. Tao et al. [142] in their algorithm WARM employed weighted support instead of minimum support threshold. Based on the significance of items, a weight is assigned to each item and only those items are considered further that satisfy the predefined weight threshold. However, assigning proper weights to each item adds to complexity of the algorithm. Wang et al. [153] in their algorithm Adaptive Apriori, pushed some support constraints on the itemsets. The lowest minimum support is considered, in case two or more constraints are applied on the itemsets. Maintaining the ordering of items even at run time becomes a tedious affair for the algorithm. DCS Apriori developed by Selvi and Tamilarasi [129] uses two support thresholds: Dynamic and Collective. Using the Dynamic support count, significant rare items are retained and the items that do not satisfy the Collective support are removed. Although the algorithm is independent of the user-defined threshold, it fails to produce the complete set of rare items. Sadhasivam and Tamilarasi [127] proposed Automated Apriori Rare that automatically assigns the support thresholds to items to derive the frequent as well as rare itemsets. The algorithm employs the strategy of MS-Apriori to extract the rare itemsets and Apriori to derive the frequent itemsets. The algorithm has the advantage of operating in parallel but misses out some significant rare itemsets.

## Extensions of FP-growth

Apriori-based techniques prove to be inefficient while mining the rare itemsets since there will be a rapid escalation in the number of candidate itemsets as the rare items are also retained during the itemset generation phase. To overcome the shortcomings of Apriori strategy, some rare pattern mining techniques have adopted the concept used by FP-Growth.

The first in this list is the CFP-Growth algorithm that extends FP-Growth using “multiple minimum support framework” to mine the rare itemsets. The algorithm stores the information about the itemsets in a tree structure called Minimum Item Supports (MIS) tree. The algorithm proved to be highly scalable, even though the tree construction phase is a bit costlier. Kiran and Reddy [75] further extended the CFP-Growth algorithm in their proposed approach, Maximum Constraint based Conditional Frequent Pattern Growth (MCCFP). The authors adopted the maximum constraint model to assign individual Minimum Item Support (MIS) values to the items. The algorithm proved to be little expensive than CFP-Growth algorithm due to an additional step of item pruning. The Multiple Minimum Support using Maximum Constraints (MSFP) algorithm proposed by Elgaml et al. [36] also employed the maximum constraint model. The

**Table 1** Datasets used

Datasets	Number of transactions	Number of items	Average transaction Size	Type
Mushroom	8124	119	23	Dense
Connect-4	67,557	129	43	Dense
Gazelle	59,602	267	10	Sparse
Retail	88,162	16,470	10	Sparse

algorithm proceeds to generate the MIS trees for only those itemsets that fulfill the predefined MIS value. The algorithm is faster than the previous approaches but fails to generate the complete set of rare items.

RP-Tree algorithm developed by Tsang et al. [146] finds the rare itemsets using a single minimum support threshold. It is an extension of FP-Growth algorithm that mines the rare-item itemsets. The algorithm takes into account only those transactions that possess minimum one rare item. The algorithm is highly efficient than other rare pattern mining algorithms in terms of execution time, but fails to generate the complete set of rare items. RP-Tree is further extended using multiple support thresholds by Bhatt and Patel [19] using their Maximum Constraint Based Rare Pattern Tree (MCRP) algorithm. The algorithm proved to be highly efficient due to the avoidance of costly pruning steps but again fails to produce the complete set of rare items. Gupta et al. [44] mines the minimally infrequent itemsets using another extension of FP-Growth called Inverse FP-tree (IFP-Tree) algorithm. To generate the minimal infrequent itemsets, it makes use of projected and residual trees. The residual tree is used to store the entire database except the removed items while projected tree is used to retain only the frequent items. Usage of residual trees reduces the computational complexity to great extent. The algorithm, however, fails to show appreciable performance in case of smaller dense datasets.

## Major research challenges for rare pattern mining techniques

The area of rare pattern mining has been extensively accepted and adopted by the research community. The previous section discussed the current status and development of the area of rare pattern mining. However, the existing techniques suffer from certain gaps and drawbacks that need to be resolved for efficiently handling the problem of rare pattern extraction. This section discusses some challenging issues encountered by the existent rare pattern mining techniques.

Mining rare patterns from the databases is not an easy task and may prove to be challenging at many instances. We performed several experiments on some of the widely referenced rare pattern mining techniques using different thresholds and came up with some of the crucial challenges faced by the

rare pattern mining community. The experimental analysis has been carried out to discover the issues associated with existing rare pattern mining techniques.

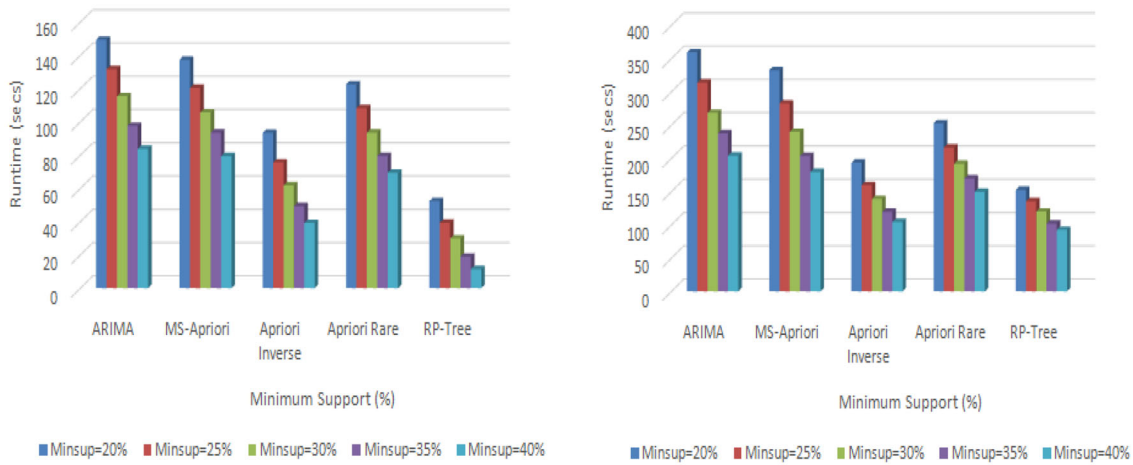
## Mining rare patterns from databases with different data characteristics

Real-life databases comprise of data having different data characteristics. The data can be frequent and dense or huge and sparse depending upon the type of application. The sparse databases contain lesser number of these frequently occurring items as compared to the dense databases. On removing the less frequent items from the databases, it is evident that only the data with the characteristics of a dense dataset will be retained. This further indicates that the rare items represent the characteristics of a sparse dataset. Most rare pattern mining techniques work well with dense datasets having many frequently occurring items but fail to handle the sparse datasets. The rare pattern mining techniques must be designed in a way that they can handle both the dense and sparse datasets efficiently.

## Experimental analysis

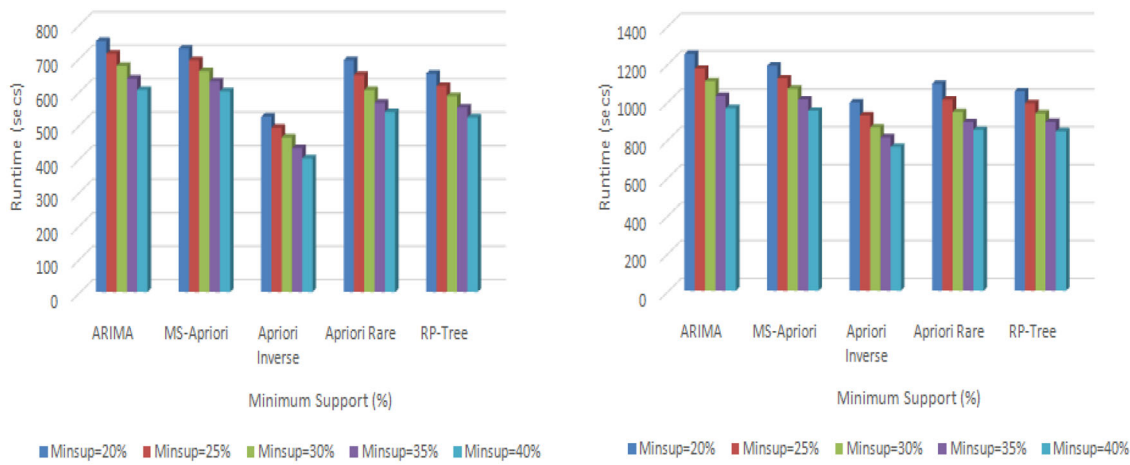
The performance of existing rare pattern mining techniques on dense and sparse datasets is evaluated using various real-life datasets. Dense datasets Mushroom and Connect-4 and sparse datasets Gazelle and Retail have been used for experimental evaluation that were obtained from UCI Machine Learning Repository. The dataset characteristics are given in Table 1. Five well-known rare pattern mining techniques have been considered: ARIMA, MS-Apriori, Apriori Inverse, Apriori Rare and RP-Tree. The considered minimum support value, Minsup starts from 20%, gradually increasing to 40%. For MS-Apriori, the value of  $\beta$  is taken as 0.1 and for Apriori Inverse, the maximum support value Maxsup is taken as 60%.

Figure 3a–d depicts the execution time invested by the rare pattern algorithms while extracting rare patterns from the four different datasets. As can be observed from the figures, the execution time required for sparse datasets Gazelle and Retail is quite high compared to dense datasets Mushroom and Connect-4. The comparative analysis given in Fig. 3e for the execution times of Gazelle and Connect-4 illustrates



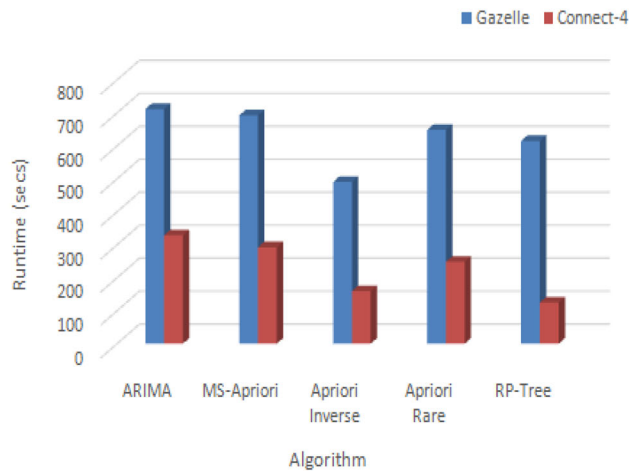
(a) Runtime on Mushroom

(b) Runtime on Connect-4



(c) Runtime on Gazelle

(d) Runtime on Retail



(e) Comparative Analysis of Execution Time

Fig. 3 Experimental evaluation on dense and sparse datasets

this fact. The performances of Apriori-based approaches are better than FP-Growth based approaches in case of sparse datasets. It is noteworthy that the execution time of RP-Tree is better than ARIMA, Apriori Rare and MS-Apriori even in sparse datasets, as it considers only a subset of rare itemsets called rare-item itemsets.

However, Apriori-based approaches like ARIMA, Apriori Inverse, Apriori Rare and MS-Apriori constructs hashing tree for the candidates are generated, in order to match and update their counts while scanning a transaction containing that particular candidate. This adds to the computation complexity of these algorithms. FP-Growth based algorithm RP-Tree, on the other hand fails to compress the FP-Tree generated effectively. Since the number of frequently occurring items is less in case of sparse datasets, sharing of nodes between items in the tree will also be less resulting in a big and bushy FP-Tree.

### Mining rare patterns from advanced data types

The problem of rare pattern mining can be extended to handle various advanced data types. The basic rare pattern mining techniques, however, directly cannot address the variations demanded by these advanced data types. Modifications of existing techniques need to be developed for effectively handling the advanced data types.

### Mining rare patterns from sequential databases

Sequential databases are a collection of events or elements that are ordered in some sequence and reported either using the time constraint or without using the same. The problem of sequential pattern mining is very much similar to that of frequent or rare pattern mining. The only difference is that a temporal ordering or a sequence is maintained for the items of a transaction in case of sequential pattern mining. Mining sequential patterns is a gruesome task that involves the exploration of a huge number of sequential patterns generated. The situation is primarily worse due to the repetition of items in the sequence and the generation of an enormous number of candidate sequences. Due to the growing length of sequences, the mining techniques need to perform multiple database scans.

Let us consider an example, to better comprehend the problem. A sequential database having a sequence of items with their respective IDs is shown in Table 2. There may exist some patterns within the sequence of items. For instance, if the minimum support threshold is assumed to be 2, then a frequent pattern in the sequence will be  $\langle (ac)d \rangle$  appearing in transactions 1 and 3 and rest of the patterns will be rare. This crucial issue has not been explored extensively by rare pattern mining algorithms and very limited attempt can be found in the literature for mining the sequential rare patterns.

**Table 2** An example of sequential database

ID	Sequence
1	$\langle b(\underline{ace})(bd)a(ef) \rangle$
2	$\langle (bd)c(ab)(de) \rangle$
3	$\langle (cf)(\underline{ac})(ef)\underline{db} \rangle$
4	$\langle fg(bf)ada \rangle$
5	$\langle e(dc)fb(a) \rangle$

### Experimental analysis

The rare pattern mining techniques do not take into account the sequential ordering of the items in the sequential database. We carried out experiments on synthetic sequential datasets to analyze the applicability of general rare pattern mining techniques for sequential rare pattern mining. The synthetic datasets were generated using the sequential dataset generation techniques described in [133]. Two synthetic datasets were used: C10-T5-S4-I1.25 and C10-T5-S4-I2.5 where C is the average number of transactions per data sequence, T is the average number of items per transaction, S is the average length of maximal potentially frequent sequences and I is the average size of itemsets in maximal potentially frequent sequences. The number of maximal potentially frequent sequences has been set to 500, the number of maximal potentially frequent itemsets has been set to 2500, the number of items has been set to 1000 and number of data sequences has been set to 10,000. The execution times invested by the rare pattern mining algorithms on these synthetic sequential datasets is shown in Fig. 4a, b.

From the obtained results it has been observed that the patterns were generated based on their frequency of occurrence and no sequential ordering of patterns has been maintained. The patterns generated can be regarded as unproductive as sequential relationship between items was ignored. Thus, suitable modifications of existing rare pattern mining techniques are needed that can handle the sequential datasets efficiently, taking into account the sequential relationships of items as well.

### Mining rare patterns from time-series or spatiotemporal databases

Time series databases involve time series data that are sequence of equally spaced points over time. The applications of time-series data range from weather forecasting and signal processing to astronomy and pattern recognition. Seismic time-series waveform data are given in Fig. 5a. The diamonds in the graph represent some seismic activity while the squares denote the time when earthquake happened. By finding hidden useful rare patterns from the data, one can predict the occurrences of earthquake. Searching for patterns in time

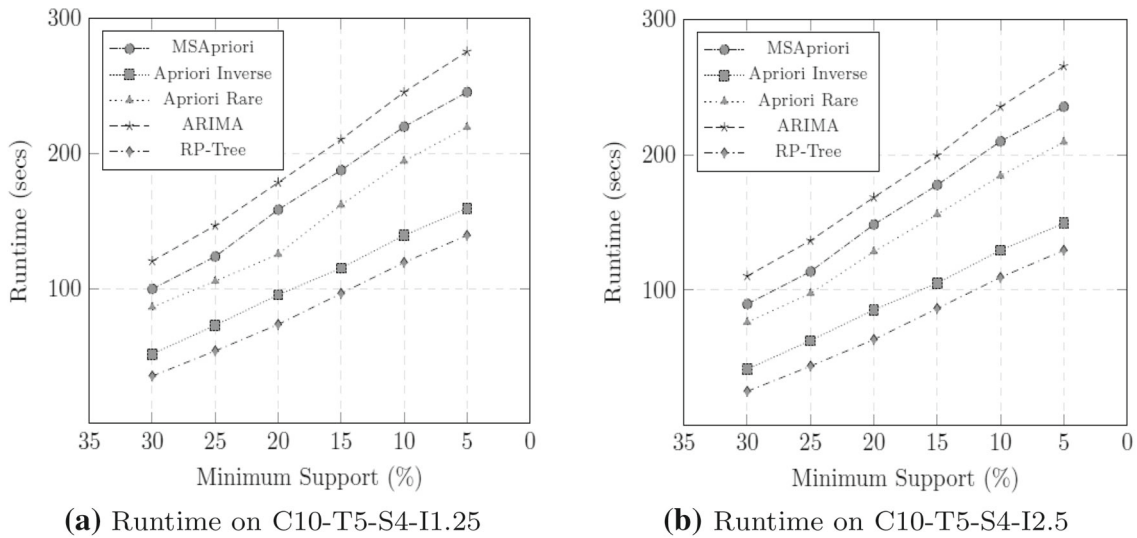


Fig. 4 Evaluation of rare pattern mining algorithms on sequential datasets

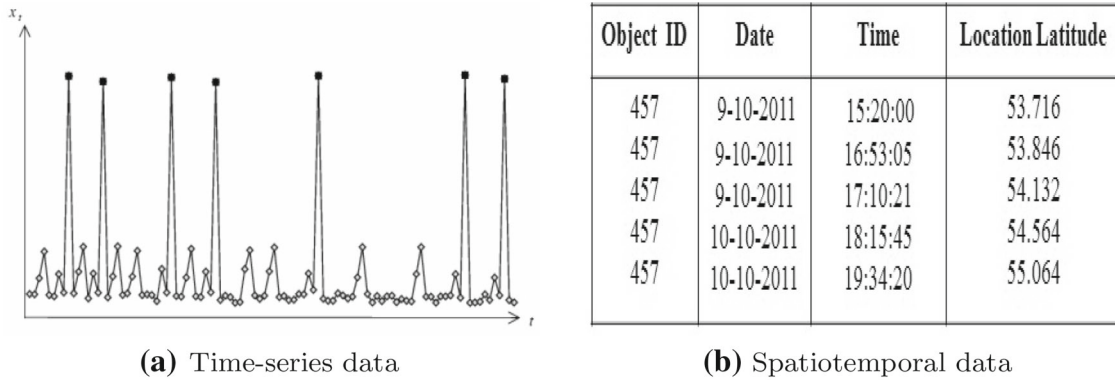


Fig. 5 Examples of time-series and spatiotemporal data

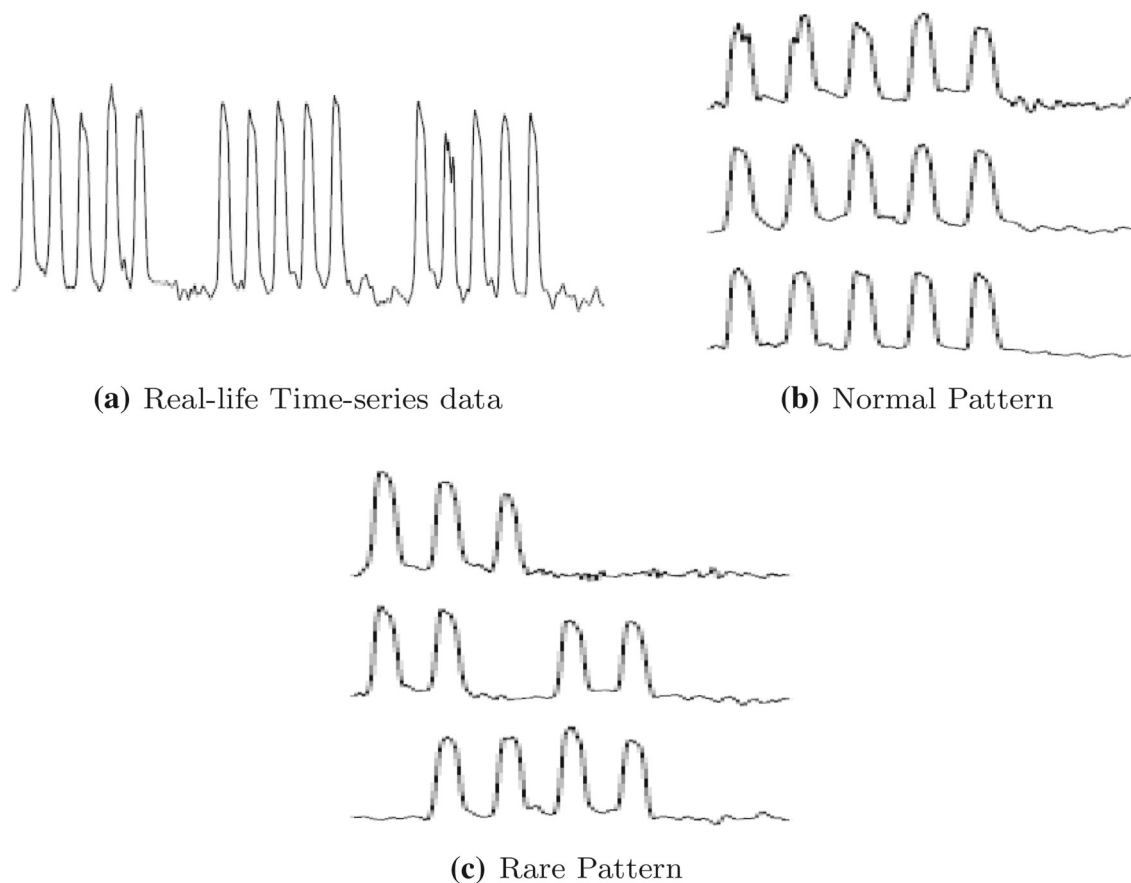
series databases is both an intriguing as well as a challenging issue. Pattern mining techniques have attempted to discover different types of periodic patterns from time series data.

Spatial databases basically comprises space-related data for instance, maps, medical image and remote sensing data. Spatiotemporal databases, on the other hand, involve a time factor for the spatial data. Spatiotemporal data are either ID based obtained from GPS or location based obtained from sensors. A sample of spatiotemporal data is shown in Fig. 5b. The sample is of a moving object data with varying sampling rate. Extracting knowledge and patterns from such data is not an easy task. Spatiotemporal data are highly complex in nature and due to the high computational cost, mining patterns from such databases become a costly affair. Mining time series and spatiotemporal databases, therefore, needs utmost attention.

### Experimental analysis

The existing rare pattern mining techniques cannot be directly applied to time-series data. Therefore, instead of performing experiments we analyzed a well-known benchmark dataset. To recognize the importance of rare pattern mining on time-series data, we considered a real-life time series dataset used by Van et al. [148]. The dataset contains a Dutch research facility’s power demand for the year 1997. The data have 35,040 points sampled over an average period of 15 min. It represents the power demand generated after every 15 min for each day of every month after for the year 1997. The interesting point about this dataset is that despite greater uniformity as shown in Fig. 6a, there are certain regions of irregularities that correspond to some rare activity. The usual and expected pattern of 5 weekday peaks followed by a flat weekend is violated due to the fact that there are certain weeks where one or more days are holidays.





**Fig. 6** Real-life time series data analysis

Figure 6b represents normal pattern of power demand for 3 weeks with no holidays. The 5 peaks in the figure denote 5 weekdays followed by a flat weekend. Figure 6c, on the other hand, represents a surprising and abnormal pattern different from the normal trend as the concerned weeks contain holidays. Such information indicates the that amount of power demand is less not only at weekends but also on holidays. The analysis thus establishes the fact that significant rare patterns may exist in time-series data, identification of which may prove to be beneficial for the research community.

### Mining rare patterns from graph databases

Structured graphs nowadays are being used for representing various types of data including chemical, biological and XML data as well as software program traces to name a few. Graph databases store and represent the data in the form of graph structures with nodes and edges. The main concept behind graph databases is the edge relationship between data items that represent how these items are related to each other. An example of the data stored in a graph database is shown in Fig. 7. The example contains the information of some employees working in a company named Robert

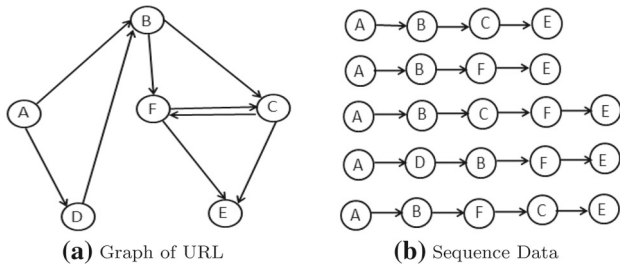
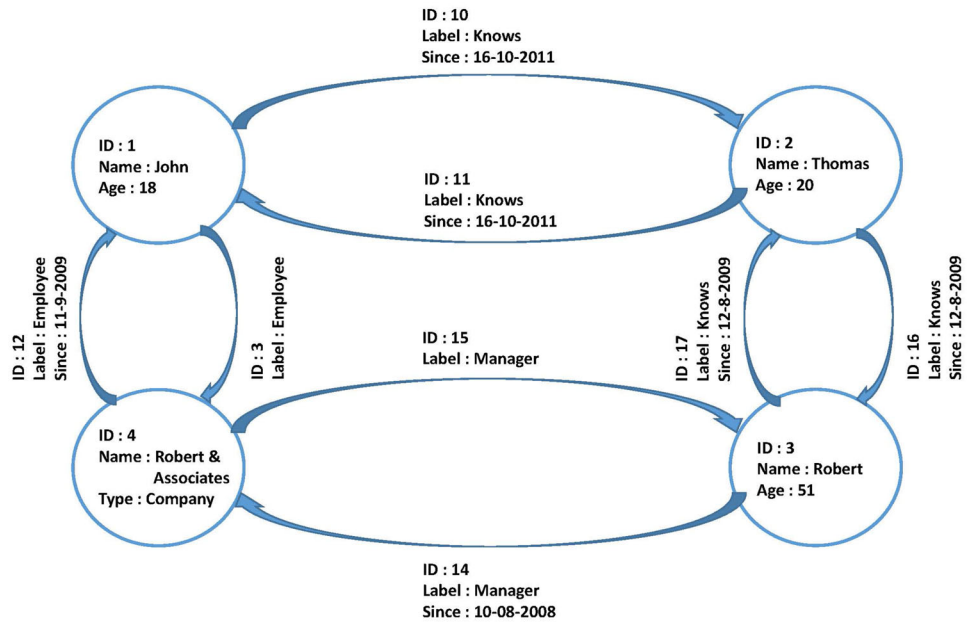
& Associates. Each node in the graph contains individual information about the employees like their id, name and age while the edges represent the relationships between them. For instance, an information conveyed from the graph will be:- “Thomas having id 2 and age 20, knows Robert having id 3 and age 50 since 12-8-2009, who is the manager of the company Robert & Associates.”

The graph databases may contain many significant patterns applicable to several domains. Mining of rare structural patterns from these databases is both an essential and intricate task that comes up with numerous challenges. First, graphs are very complicated structures and mining rare patterns from them itself is a challenging issue. Second, the most crucial hurdle while mining graph databases is their massive size. The graph databases are so large and complicated that they mostly cannot fit in main memory. This demands for scalable graph mining techniques that can employ secondary memory during the mining process.

### Experimental analysis

As the current rare pattern mining techniques are not suitable for graph mining, we decided to perform data analysis using

**Fig. 7** An example of data stored in graph database



**Fig. 8** Graph data analysis

an existing frequent graph pattern mining technique. For data analysis, we have used a real-life web browsing dataset. The dataset contains one-day log access history of a WWW site known as Achara NAVI of Recruit Co. Ltd. We used the freely available implementation of a well-known graph pattern mining algorithm called Apriori-based Graph Mining (AGM) [66] for analyzing the data. Each line of the log text file contains information about the user’s IP address, visited URL and access points. The access log was converted into a set of transactions by removing the access points and IP addresses. Each access history forms a transaction. The access log file contains nearly 8700 URLs with their associated links. These links were further transformed into nodes using the graph representation technique described in [67]. For ease of understanding, the URL’s were mapped into alphabets representing the node labels. The users visit the URL’s with the help of the hyperlinks. The graph shown in Fig. 8a illustrates the URLs along with their associated hyperlinks. The sequence of access is given in Fig. 8b.

From the figures it can be observed that for reaching E from A, the mostly used path is via B then from B via some

other URL’s like C or F. The frequent sequence in this case is from A to E via B and C or via B and F. There is another sequence starting from A to D then reaching E via B and F. Even though this sequence has been rarely followed, it suggests that if a link had been present between URL’s D and E, then the clients could have directly accessed E from A via D. Such a sequence would have been more accessible and uncomplicated. Thus identifying such rare sequences in graph can aid in faster and smoother access of the URLs. This analysis, therefore, justifies the importance of identifying rare patterns from graph databases.

### Mining rare patterns from large and high dimensional databases

With emerging technology, there is a rapid growth in the size of real world databases. The pattern mining techniques are heavily dependent on main memory and thus become incompatible when it comes to handling large databases. Similarly, working with high dimensional bioinformatics data like microarray and gene expression data is a challenging issue as the datasets contain hundreds and thousands of columns.

The general pattern mining algorithms depend on the row length of tables to a great extent. With increase in row length or the number of columns, the combination of items become exponential, which poses great difficulty in front of the pattern mining techniques. Due to the retainment of rare itemsets, this situation particularly becomes worse in case of rare pattern mining techniques. Thus efficient rare pattern mining techniques are needed that can scale well with increasing database size or dimension.

## Experimental analysis

To test the scalability of rare pattern mining techniques, we performed several experiments using synthetic datasets. The synthetic datasets were generated using the data generation process described in [8]. Two synthetic datasets, T10I4 and T20I10, were used varying the number of transactions. The first has an average transaction length of 10 items, average frequent itemset size of 4 items and the number of distinct items 1 K while the second one have a similar number of distinct items but an average transaction length of 20 and average frequent itemset size of 10 items. The size of the database has been varied from 100 to 1000 K and a Minsup of 10% has been used. To gauge performances in terms of scalability, we considered a main memory limitation of 256 MB.

Figure 9a, c shows an obvious increase in the execution time of the algorithms with increase in the number of transactions for the two datasets. From Fig. 9b, it can be observed that the algorithms managed to complete their execution within the memory limitation of 256 MB. However, for dataset T20I10, the execution of algorithms ARIMA, MS-Apriori and Apriori Rare could not be completed within the restricted amount of memory as illustrated in Fig. 9d. Therefore, it can be concluded that for a memory limitation of 256 MB, these algorithms would have failed as their memory usage has increased beyond the available main memory size.

To assess performance with respect to memory usage, we calculated the memory utilization factors for each algorithm. Memory utilization factor is computed as:- available main memory/memory usage of algorithms [4]. Thus, memory utilization factor for an algorithm decreases with increase in the amount of memory usage. From Fig. 9e, it can be witnessed that the performance of algorithms degrade with increase in memory usage or decrease in memory utilization factor.

## Mining rare patterns from incremental data

Real-world applications undergo tremendous modifications due to continuous addition and deletion of data. Handling dynamic databases is a gruesome task for pattern mining techniques as the data are continuously updated, generating new rules and invalidating the existing ones. Pattern mining techniques invest huge amount of time, processing the newly updated database.

Let us consider an example to better understand the problem. Example of an incremental database is shown in Table 3. Database D in Table 3a represents the original database while Table 3b represents the updated database. For a support threshold of 3, the frequent items in database D are:- a, b and e whereas c, f and d are the rare items. In the updated database D', D- represents the transactions that were deleted while D+ represents the newly added items. After the update occurs, the frequencies or support values of the items will

change. For instance, in the newly updated database, a, b and d will be the frequent items while c, f and e will be the rare items. Thus item d which was initially rare, has now become frequent due to the change in its support value and item e which was frequent in the original database, has now become rare. Thus, in order to overcome such variations in support count of the items, advanced techniques are needed that operate only on the updated or incremental part of the database instead of processing it from scratch.

In case of rare pattern mining, the incremental update of data may make some rare itemsets frequent and some previous frequent itemsets rare due to change in support count values of these items. These may invalidate the entire set of rare association rules generated. The existing rare pattern mining techniques assume the transactional data to be static and operate without considering the dynamic nature of databases. There is a need for expansion of these techniques to work with dynamic databases as well.

## Experimental analysis

For experimental evaluation, we have used the same synthetic dataset T10I4, that was used in the previous experiment. We fixed the size of the initial database to 10,000 rows. To test the performance of the rare pattern mining techniques on incremental databases, we kept adding different increments and the experiments were performed on varying increment sizes. The increment sizes were varied from 500 transactions, that constitutes 5% of the original database to 5000 transactions, constituting 50% of the original database. The Minsup value has been set to 5%.

The execution times of the algorithms on the original database and different increment sizes are depicted in Fig. 10a, b, respectively. Figure 10c on the other hand, illustrates the runtime invested by the algorithms on updated database with respect to the original database. From the figures, we can observe that the performance of the algorithms degrade upon addition of the increments. The execution time spent on the updated database is quite higher than the original database and it kept on increasing with smaller increments to very large increments. This is quite obvious since the algorithms run from scratch when new transactions are added to the original data. This clearly establishes the inefficiency of the rare pattern mining techniques in handling incremental data and the need of competent techniques for handling the same with greater adaptability.

## Mining rare patterns from data streams

The online applications in various domains, generate enormous volume of data streams at a rapid rate. In case of data streams, the flow of data is continuous and varying unlike traditional databases. The storage of data streams is a

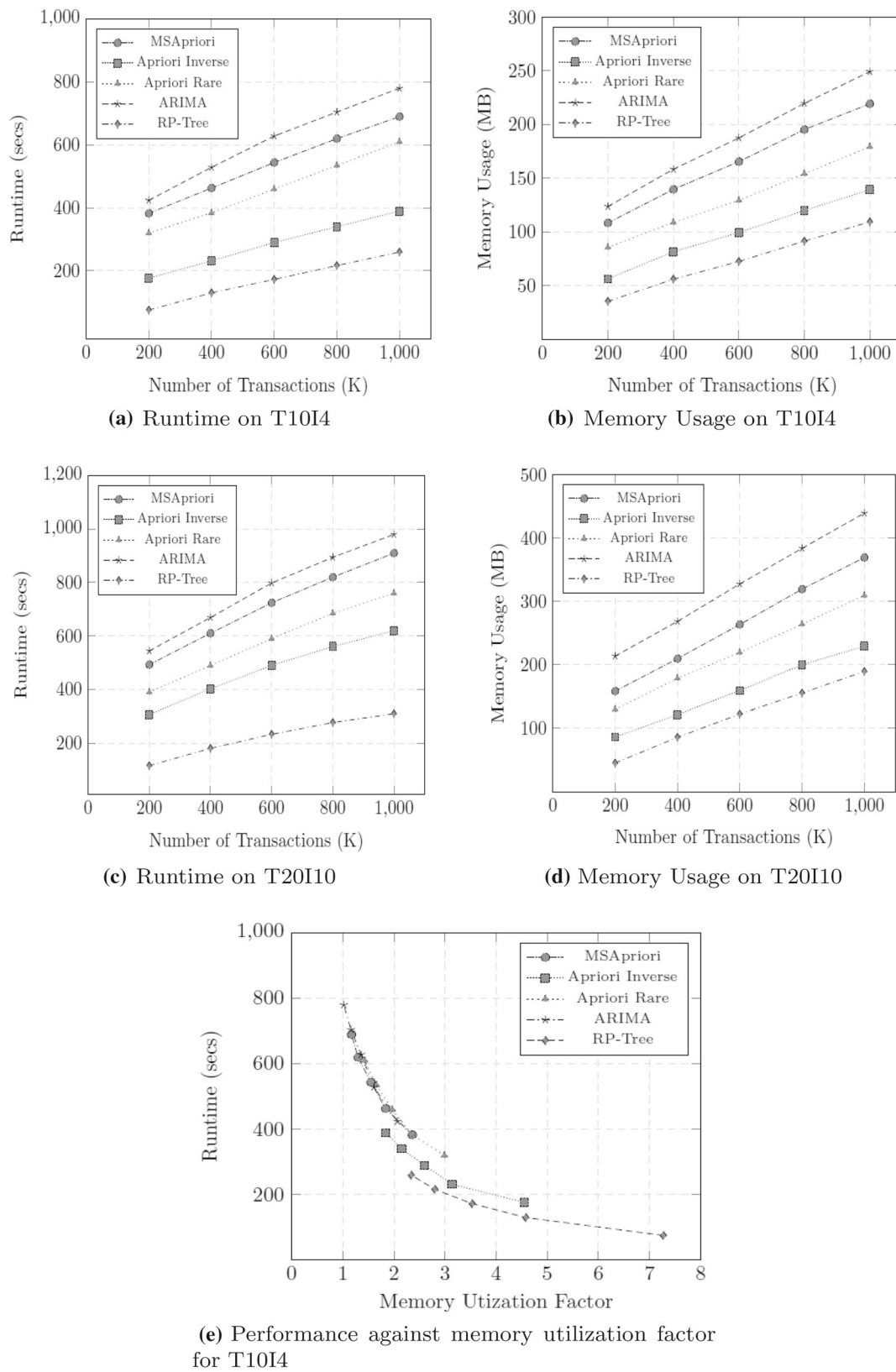


Fig. 9 Scalability evaluation of rare pattern mining algorithms

**Table 3** Example of incremental database

Tid	Items	Tid	Items
(a) Original database		(b) Updated database	
D		D'	
1	a, b, c	D–	1 a,b,c
2	a, b, e	2	a, b, e
3	a, b, d, f	3	a, b, d, f
4	a, b	4	a, b
5	c, d, e	5	c, d, e
6	a, b, e, f	6	a, b, e, f
		D+	7 a,b,d
		8	a, b, c, d

gruesome task considering its tremendous volume and high speed. Moreover, scanning the entire data stream multiple times makes the rare pattern mining technique inefficient and incompetent. For online extraction of rare patterns, single pass rare pattern mining techniques are needed. With the passage of time, the frequency counts of itemsets may change making them frequent or rare, an issue known as concept drift. The rare pattern mining techniques must be capable of handling the concept drift problem efficiently.

### Experimental analysis

The significance of rare patterns in data streams is examined using a real-life stream dataset called RSS feed. We collected sports and entertainment news stories from the database. The processing has been done using WordNet and the data has been divided into 14 streams, each stream representing data for a duration of 24 h. The streams obtained after preprocessing are ordered sequence of entertainment and sports stories generated each day. To obtain the ground truth or the number of rare patterns in the stream sequence, the data has been analyzed manually. The graph given in Fig. 11 demonstrates the number of rare items identified each day.

The rare items represent the stories whose words are not frequently repeated in subsequent stories for a period of 24 h. These infrequent stories, however, have been found contain some significant sports and entertainment information that substantiate the implication of rare pattern mining over data streams.

### Frequent vs rare pattern mining: a comparison

Over the years, there has been a considerable development of pattern mining techniques deriving patterns from different kinds of databases. Even since its inception, frequent patterns

were the only concern for the pattern mining community. The rare patterns were considered to be of least importance that give no valuable information to the users as such. With the advent of technology, rare patterns have proved to be of vital importance in many domains. The area of frequent pattern mining has been extensively studied, however many developments have been made in the field of rare pattern mining as well. This section elicits a comparative study between these two areas. The comparison is provided to give convincing facts as viable future directions for the researchers in the area of rare pattern mining.

### Issues handled by frequent and rare pattern mining techniques

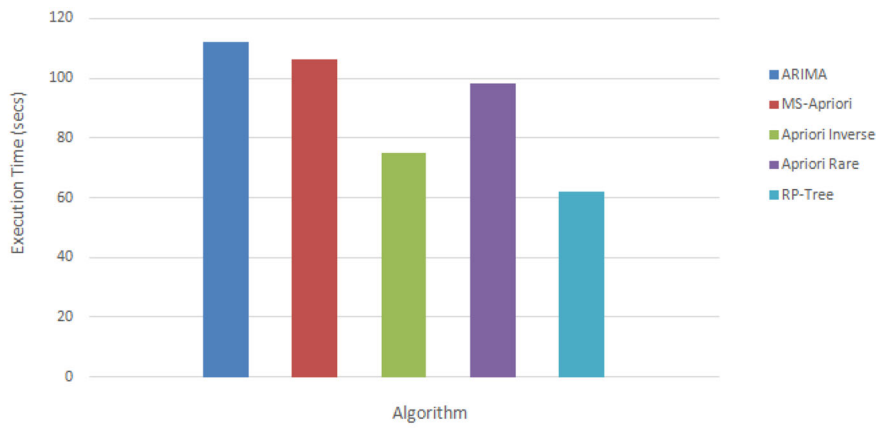
Many endeavors from frequent pattern mining techniques can be seen in the literature handling the above discussed challenges. Table 4 elucidates some of the articles on frequent pattern mining handling the various issues. Only the relevant and recent articles published in the area of frequent pattern mining have been considered. The articles handling the respective issues are only included for comparisons excluding surveys. To better understand the status of frequent pattern mining techniques under various issues, a graphical analysis is given in Fig. 12.

From the trends in the graph, it can be concluded that the highest amount of research in the field of frequent pattern mining has been carried out for mining the sequential patterns followed by mining frequent patterns from data streams. However, only a limited attempt has been made to mine frequent patterns from databases having different data characteristics.

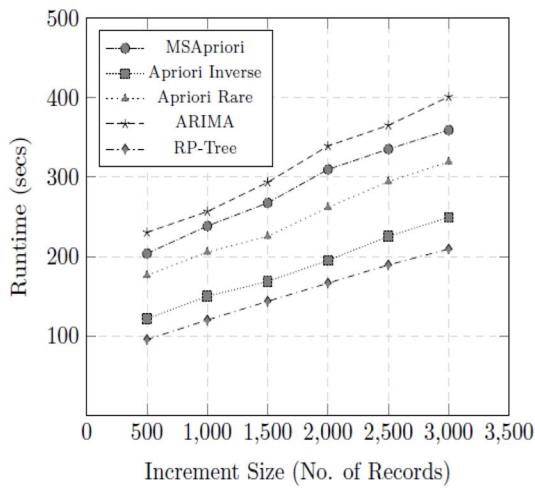
Rare pattern mining, being a new and emerging area, has attempted only few of the pattern mining issues. The various articles published in the area of rare pattern mining handling different issues are given in Table 5. The table includes only articles handling a particular issue excluding the review articles.

### Comparison between frequent and rare pattern mining techniques

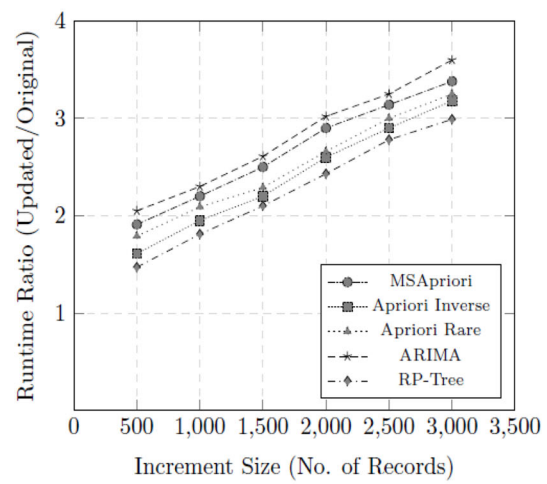
To make the study more convincing, a comparison between frequent and rare pattern mining areas based on the amount of research carried out, a graphical analysis is provided in Fig. 13. The figure illustrates that only a limited attempt has been made by rare pattern mining techniques for mining patterns from data streams, graph databases and sequential patterns as compared to frequent pattern mining techniques. There are still no endeavors towards mining rare patterns from incremental and large databases as well as databases with different data characteristics.



(a) Runtime on Original Dataset T10I410K



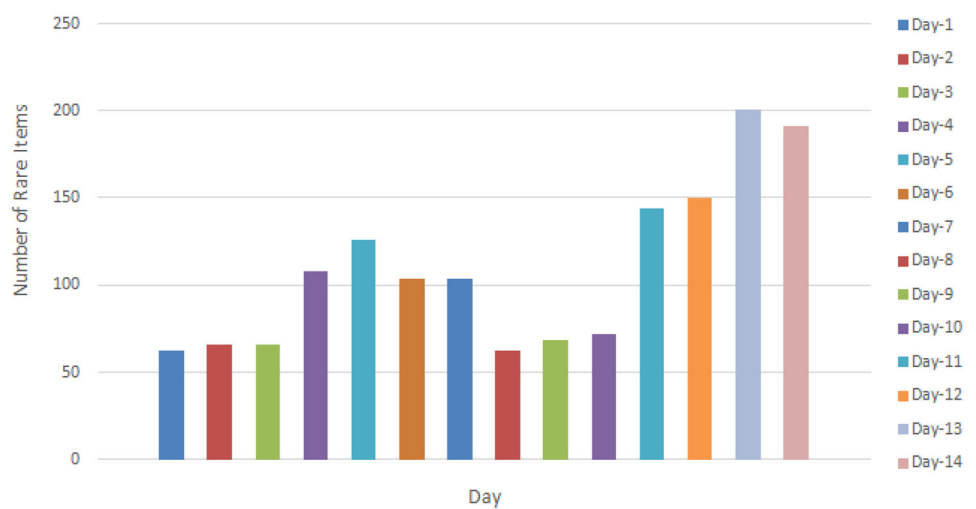
(b) Runtime on Updated Dataset



(c) Runtime Ratio of Updated Database to Original Database

Fig. 10 Evaluation of rare pattern mining algorithms on incremental datasets

Fig. 11 Number of rare items in RSS feed data



**Table 4** Issues handled by frequent pattern mining techniques

Issues	Articles handling the issues	No. of articles
Mining data with different data characteristics	[6,7,16,81,110,114,115]	7
Mining sequential patterns	[1,9,12,21,23,26,27,32,39–41,49,51,54,62,63,72,73,79,82,86,102–104,106,112,116–118,121,122,130,133,147,152,161,163,164]	38
Mining time-series and spatiotemporal databases	[17,20,35,43,47,48,60,61,64,64,76,78,100,108,126,131,134,155,162]	19
Mining graph databases	[18,25,57,66,80,92,113,136,149,150,154,158–160,171,174,175]	17
Mining large and high dimensional databases	[4,5,14,15,94,97,110,111,119,139,172]	11
Handling incremental data	[10,11,13,22,28–31,37,55,65,71,89,95,107,137,140,157,167]	19
Handling data streams	[24,33,34,42,69,70,84,87,88,90,91,93,98,99,101,105,109,125,138,141,143,145,151,156,165,168,169]	27

## Future directions for rare pattern mining

Rare pattern mining is a relatively less explored area than frequent pattern mining. The growing urge for rare patterns in various domains indicates that the field of rare pattern mining is emerging extensively and there is much room for expansion. This section discusses some future prospects for the area of rare pattern mining.

### 1. Mining sparse data and datasets with long patterns

The existing rare pattern mining techniques can very well handle the dense datasets containing a group of frequent items. As discussed in “Mining rare patterns from databases with different data characteristics”, rare pattern mining techniques fail to handle sparse data and datasets with long patterns. The rare pattern mining techniques are primarily based on Apriori or FP-Growth like approaches. Both Apriori and FP-Growth approaches fail to handle sparse data and datasets with long patterns efficiently. Thus there is a growing urge for rare pattern mining techniques that can efficiently handle sparse data and datasets with long patterns.

A possible solution could be to use array-based or queue-based implementation instead of a tree-based implementation as the performance of tree data structure is not substantial in case of sparse datasets.

### 2. Scalable rare pattern mining algorithms

Scalability of algorithms is a prime issue to be considered during the mining of rare patterns. Although the frequent pattern mining techniques have extensively handled the scalability issue, existing rare pattern mining techniques have not taken any initiative in this regard. Memory usage,

being one of the crucial issues of pattern mining, must be taken into account during the generation of rare itemsets. In order to mine both frequent and rare itemsets, the rare items need to be preserved along with the frequent one during itemset generation phase. This clearly indicates that the requirement of memory will be more, due to which rare pattern mining methods might have to go beyond physical memory and employ secondary memory in the mining process.

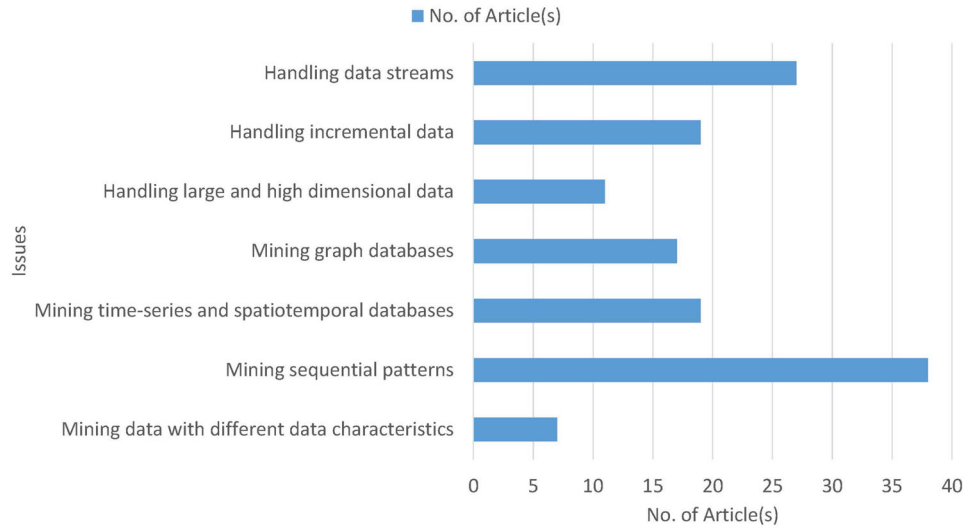
A probable approach towards solving this issue could be to employ secondary storage structures where the candidate itemsets and nodes representing the rare itemsets can be stored when they can no longer be accommodated in main memory.

### 3. Incremental rare pattern mining algorithms

Rare pattern mining techniques mostly operate on transaction databases that are static in nature. In case of static databases, there is no addition or deletion of items as such and hence they are not updated every time. Mining rare patterns from dynamic or incremental databases is a bigger challenge as the databases undergo continuous update due to the inclusion or erasure of transactions. The rare pattern mining community, although has not considered this crucial issue but generating incremental algorithms for mining rare patterns, does not seem to be an unachievable task considering the extensive efforts in the field of frequent pattern mining.

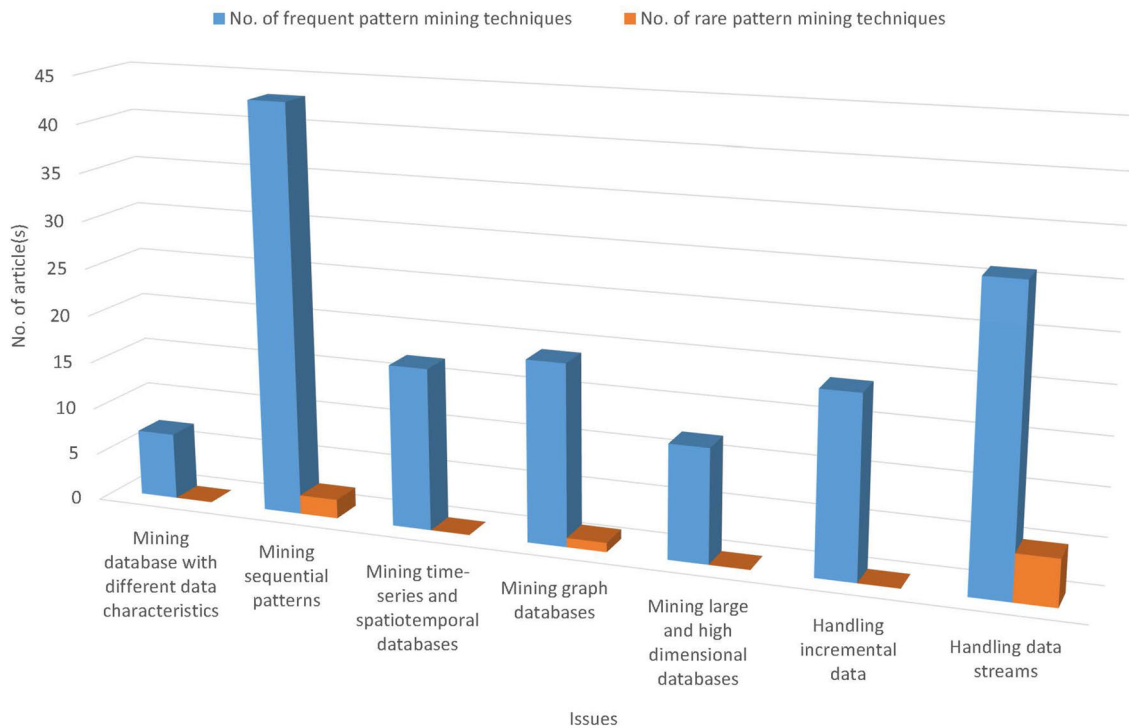
The existing rare pattern mining algorithms use tree data structures like FP-Tree or RP-Tree. Thus one can think in the direction of extending FP-Tree or RP-Tree to handle incremental datasets in the same manner as that of

**Fig. 12** Number of frequent pattern mining articles reviewed



**Table 5** Issues handled by rare pattern mining techniques

Issues	Articles handling the issues	No. of articles
Mining data with different data characteristics	Nil	–
Mining sequential patterns	[124,173]	2
Mining time-series and spatiotemporal databases	Nil	–
Mining graph databases	[170]	1
Mining large and high dimensional databases	Nil	–
Handling incremental data	Nil	–
Handling data streams	[53,58,59,83,128]	5



**Fig. 13** Comparison of frequent and rare pattern mining techniques based on articles reviewed under various issues



IFP-Tree. Another solution could be to use the strategies of FUP or Borders algorithm.

4. **Mining rare sequential patterns** Sequential pattern mining is a significant area that comes up with numerous challenges. “Mining rare patterns from sequential databases” discussed the issues involved in the extraction of rare sequential patterns from databases. Mining sequential patterns is a less explored area for the rare pattern mining community.

Recently, Zhu et al. [173] has made an attempt to mine the rare sequential patterns from document streams which are sequential in nature. The algorithm named Sequential Topic Patterns(STPs) tries to identify the abnormal activities of Internet users over document streams. A similar attempt has been made by Rahman et al. [124] to detect anomalies in SCADA logs. The mining of rare sequential patterns has not been extensively explored and there is much room for expansion in this regard. The rare pattern mining community needs to develop efficient rare sequential pattern mining techniques maintaining the temporal and sequential relationships among the rare patterns.

5. **Mining rare patterns from graph databases** Interesting patterns may appear in graph databases that make graph pattern mining an interesting concept to work on. In case of pattern mining, transactions in graph are symbolized using adjacency matrices. Many attempts can be found in the literature to mine frequent patterns from such databases. The appreciable amount of work done in the area of frequent pattern mining for graph databases advocates exploration of the same for mining rare patterns. The enormous amount of research for mining patterns from graph databases suggests that it is not an inconceivable task and can be extended for mining rare patterns with substantial effort.
6. **Mining rare patterns from time-series and spatiotemporal data** Mining patterns from time-series and spatiotemporal data have vital importance in the field of data mining. Interesting patterns may exist in such time-related databases that may be beneficial for various application domains. Rare events or patterns present in the time series or spatiotemporal databases may provide some useful information valuable for the researchers. Therefore, the rare pattern mining community must emphasize on the mining of such databases to establish rare pattern mining as a significant area in the field of data mining.

The rare periodic patterns from time-series or spatiotemporal databases are likely to deliver some significant information to the users. Thus rare pattern mining community must show tremendous efforts in discovering such beneficial patterns by pursuing the strategies used by frequent pattern mining techniques.

7. **Mining rare patterns from data streams** Extracting patterns over data streams has always been an indispensable task in the field of data mining. Rare pattern mining although being a new area has also attempted few endeavors towards the mining of rare patterns from such data. Nevertheless, there is much room for expansion in this area when it comes to mining of rare patterns from data streams. Both single pass [58,59,128] and multiple pass algorithms [53,83] have been designed for generation of rare patterns from data stream. The techniques developed, however, cannot efficiently handle the issue of concept drift. The rare pattern mining techniques can refer the existing efficient frequent pattern mining techniques to develop suitable techniques for mining rare patterns over data streams, efficiently handling the issue of concept drift in near future.

8. **Mining multilevel and multidimensional rare association rules** The transactions in the databases can be conceptually organized at different levels that calls for the extension of rare itemsets and rare association rules to multiple levels of abstraction as well as to multiple dimensions. In many applications, the data is assumed to be sparse at lower levels while at higher levels, the data is mostly dense. The rules obtained from the lower levels of abstraction are generally considered to be less interesting or not strong enough.

Srikant et al. [132] as well as Han et al. [46] suggested that for a uniform support value across different levels, higher level frequent itemsets need to be obtained first followed by their corresponding lower level itemsets. Han et al. [52] also suggested the use of different support values for different levels of abstraction. Considering all the above facts, there is a need to extend the concept of mining rare association rules to multiple levels and dimensions.

9. **Mining rare patterns using vertical data format:** Rare pattern mining techniques generally extract the patterns by mining the database in a horizontal format where one column represents the transaction id, whereas the other represents the number of items brought together in the transaction with that particular id. An alternative to this could be the mining of database in vertical data format.
10. **Mining multi-objective rare association rules:** Most of the pattern mining techniques handle the issue of extracting association rules by employing a support-confidence framework and hence considering the same to be a single objective issue. However, many studies in the literature claim that it is not enough to measure the interestingness of a rule using only their support and confidence values. This suggests the rare pattern mining community to look beyond the support-confidence framework and incorporate additional measures to judge the interestingness of a rule. Some rare association rule mining algorithms use additional parameters like comprehensibility and inter-

estingness along with support and confidence to measure the quality of a rule.

With the purpose of mining multiobjective rare association rule, only few attempts have been made in the literature. The first one by Hoque et al. [56] in their algorithm FRIMA and the second one being the RP-Tree algorithm developed by Tsang et al. [146]. In case of FRIMA, in addition to support and confidence parameters, comprehensibility and interestingness are also used to measure the quality of a rule. RP-Tree employed seven different rule interestingness measures to gauge the quality of the rare rules generated. To obtain significant interesting rare rules, there is a need to employ different rule interestingness measures, instead of sticking to the mere ‘support-confidence framework’. The rare association rule generation algorithms must extend their concept of rule quality to different interestingness parameters.

**11. Mining rare patterns from big data** With the advent of technology, tremendous volume of data is being generated by the databases, processing of which become a challenging task. Big data as the name suggests, refers to those datasets that cannot be handled by the traditional data processing techniques due to their huge size and complexity involved.

Mining rare association rules from big data can be a future aspect for the rare pattern mining community. Although rare association rule mining itself is a challenging area, nevertheless there is scope of expansion and considering its extensive growth, mining big data does not seem to be an impossible task.

## 12. Application based future perspectives

Apart from the key challenges involved in the area of pattern mining, there are still some less explored real-life applications of vital importance that need to be addressed. The rare pattern mining community must take up these applications as well with utmost importance. This section, therefore, discusses some future directions handling significant real-life applications in the area of rare pattern mining.

(a) *Mining rare patterns from high dimensional biological datasets* Bioinformatics has emerged as a significant field of research over the years. Microarray data have been widely used in various disease related research. Some abnormalities in the gene expression levels of microarray data may indicate the presence of any disorder that can be well established by the extraction of rare patterns. The area of rare pattern mining has not yet explored such biological datasets but has much room for expansion with regard to this area. However, mining such high-dimensional datasets with different characteristics is a serious issue for pattern mining algorithms

as discussed in “Mining rare patterns from large and high-dimensional databases”. The rare pattern mining community must take up the mining of biological datasets as a substantial application to work on, looking at its emerging demand and significance. It is worth anticipating rare patterns from such data that may help in identify some serious abnormalities benefiting the medical community.

(b) *Mining rare patterns for detecting network anomalies* Network anomaly detection is a crucial problem to be resolved for the data mining community. Extraction of these anomalous patterns resulting from fraudulent activities or system failure is a serious issue to look upon. Rahman et al. [123] made an initial attempt towards it by extracting rare patterns from wireless connection records. The Apriori-based algorithm developed by them obtains the rare patterns to identify the anomalous records. The authors then proceed to find the anomalous patterns in SCADA logs using rare pattern mining in [124]. The rare pattern mining community needs to extend its endeavors towards anomaly detection taking into consideration its tremendous implications.

(c) *Mining rare patterns for identifying adverse drug reactions* The rare patterns present in electronic patient database might provide significant indications about adverse drug reactions. This application domain can be a fruitful future perspective for the area of rare pattern mining. [68] identified the causal relationships between drugs and its adverse reactions using an interestingness measure. Feldman et al. [38] employed similar strategies to mine the rare patterns for identifying drug reactions. These limited attempts, however, are not sufficient for such crucial research applications and demand much more attention.

(d) *Detecting suspicious behavior in web applications* Suspicious activity may occur that needs to be identified for secure web operations. Identification of rare patterns that correspond to fraudulent activities could assist in web application security. With such a notion, Adda et al. [3] developed a system for detecting and analyzing suspicious web applications. Significant rare pattern mining techniques encompassing such applications are much desired.

## Conclusion

Over the years there has been extensive research in the field of pattern mining. Pattern mining techniques have single handedly considered the mining of only frequent patterns, neglecting the rare ones. Recent studies illustrate the sig-

nificance of rare patterns in a wide range of application areas. From “Significance of rare patterns and rare association rules”, the relevance of these substantial patterns in various domains can be recognized. The research community has taken a step forward towards the exploration of these momentous patterns keeping in mind their significance. However, being a novel field, the scope of expansion of rare pattern mining is abounding.

Mining rare patterns from data comes up with numerous challenges. The experimental analysis provided in “Major research challenges for rare pattern mining techniques” illustrates the various issues involved in rare pattern mining. From the comparative study given in “Frequent vs rare pattern mining: a comparison”, it can be concluded that only a small amount of work is carried out for mining the rare patterns from data streams and sequential databases. However, significant quantities of work have been done for the same in case of frequent pattern mining. Furthermore, some issues like mining rare patterns from large and high dimensional databases, databases with different data characteristics, incremental databases, time-series and spatio-temporal databases are still untouched. The comparison will let the researchers gain an insight into the current status of rare pattern mining as well as gauge its future scope.

This article also attempts to provide some significant ideas as feasible future directions for the rare pattern mining community in “Future directions for rare pattern mining”. Through this brief overview, the article makes an effort to provide some practicable directions to the researchers looking for some viable future perspectives to work on.

**Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

## References

- Adam O, Abdullah Z, Ngah A, Mokhtar K, Ahmad WMAW, Herawan T, Ahmad N, Deris MM, Hamdan AR, Abawajy JH (2016) Incspade: an incremental sequential pattern mining algorithm based on spade property. *Advances in machine learning and signal processing*. Springer, Berlin, pp 81–92
- Adda M, Wu L, Feng Y (2007) Rare itemset mining. In: *Machine learning and applications, 2007. ICMLA 2007. Sixth international conference on*, IEEE, pp 73–80
- Adda M, Wu L, White S, Feng Y (2012) Pattern detection with rare item-set mining. *arXiv preprint arXiv:1209.3089*
- Adnan M, Alhaji R (2009) Drfp-tree: disk-resident frequent pattern tree. *Appl Intell* 30(2):84–97
- Adnan M, Alhaji R (2011) A bounded and adaptive memory-based approach to mine frequent patterns from very large databases.

- IEEE *Trans Syst Man Cybern Part B (Cybernetics)* 41(1):154–172
- Agarwal RC, Aggarwal CC, Prasad V (2000) Depth first generation of long patterns. In: *Proceedings of the sixth ACM SIGKDD international conference on Knowledge discovery and data mining*, ACM, pp 108–118
- Aggarwal CC (2001) Towards long pattern generation in dense databases. *ACM SIGKDD Explor Newsl* 3(1):20–26
- Agrawal R, Imieliński T, Swami A (1993) Mining association rules between sets of items in large databases. *ACM Sigmod Rec* ACM 22:207–216
- Albert-Lorincz H, Boulicaut JF (2003) Mining frequent sequential patterns under regular expressions: a highly adaptative strategy for pushing constraints. *Proc SIAM DM* 2003:316–320
- Aref WG, Elfeky MG, Elmagarmid AK (2004) Incremental, online, and merge mining of partial periodic patterns in time-series databases. *IEEE Trans Knowl Data Eng* 16(3):332–342
- Aumann Y, Feldman R, Lipshtat O, Manilla H (1999) Borders: an efficient algorithm for association generation in dynamic databases. *J Intell Inf Syst* 12(1):61–73
- Ayres J, Flannick J, Gehrke J, Yiu T (2002) Sequential pattern mining using a bitmap representation. In: *Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining*, ACM, pp 429–435
- Bakar WABWA, Abdullah ZB, Saman MYBM, Man MB, Herawan T, Hamdan AR et al (2016) Incremental-eclat model: an implementation via benchmark case study. *Advances in machine learning and signal processing*. Springer, Berlin, pp 35–46
- Baralis E, Cerquitelli T, Chiusano S (2010) A persistent hy-tree to efficiently support itemset mining on large datasets. In: *Proceedings of the 2010 ACM symposium on applied computing*, ACM, pp 1060–1064
- Baralis E, Cerquitelli T, Chiusano S, Grand A (2015) Scalable out-of-core itemset mining. *Inf Sci* 293:146–162
- Bayardo RJ Jr (1998) Efficiently mining long patterns from databases. *ACM Sigmod Rec* 27(2):85–93
- Berberidis C, Aref WG, Atallah M, Vlahavas I, Elmagarmid AK et al (2002) Multiple and partial periodicity mining in time series databases. *ECAI* 2:370–374
- Bhaskar KV, Subramanyam R, Reddy KT, Sumalatha S (2016) Top-k closed sequential graph pattern mining. *Int J Inf Eng Electron Bus* 8(4):1
- Bhatt U, Patel P (2015) A novel approach for finding rare items based on multiple minimum support framework. *Procedia Comput Sci* 57:1088–1095
- Cao H, Cheung DW, Mamoulis N (2004) Discovering partial periodic patterns in discrete data sequences. *Pacific-Asia conference on knowledge discovery and data mining*. Springer, Berlin, pp 653–658
- Capelle M, Masson C, Boulicaut JF (2002) Mining frequent sequential patterns under a similarity constraint. *International conference on intelligent data engineering and automated learning*. Springer, Berlin, pp 1–6
- Chang HY, Lin JC, Cheng ML, Huang SC (2016) A novel incremental data mining algorithm based on fp-growth for big data. In: *Networking and network applications (NaNA), 2016 international conference on*, IEEE, pp 375–378
- Chang JH (2011) Mining weighted sequential patterns in a sequence database with a time-interval weight. *Knowl Based Syst* 24(1):1–9
- Chang JH, Lee WS (2003) Finding recent frequent itemsets adaptively over online data streams. In: *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining*, ACM, pp 487–492

25. Chen C, Lin CX, Fredrikson M, Christodorescu M, Yan X, Han J (2009) Mining graph patterns efficiently via randomized summaries. *Proc VLDB Endow* 2(1):742–753
26. Chen YL, Chen SS, Hsu PY (2002) Mining hybrid sequential patterns and sequential rules. *Inf Syst* 27(5):345–362
27. Chen YL, Chiang MC, Ko MT (2003) Discovering time-interval sequential patterns in sequence databases. *Expert Syst Appl* 25(3):343–354
28. Cheng H, Yan X, Han J (2004) Incspan: incremental mining of sequential patterns in large database. In: *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*, ACM, pp 527–532
29. Cheung DW, Han J, Ng VT, Wong C (1996) Maintenance of discovered association rules in large databases: an incremental updating technique. In: *Data Engineering, 1996. Proceedings of the twelfth international conference on*, IEEE, pp 106–114
30. Cheung DWL, Lee SD, Kao B et al (1997) A general incremental technique for maintaining discovered association rules. *DASFAA* 6:185–194
31. Cheung W, Zaiane OR (2003) Incremental mining of frequent patterns without candidate generation or support constraint. In: *Database engineering and applications symposium, 2003. Proceedings. seventh international*, IEEE, pp 111–116
32. Cong S, Han J, Padua D (2005) Parallel mining of closed sequential patterns. In: *Proceedings of the eleventh ACM SIGKDD international conference on Knowledge discovery in data mining*, ACM, pp 562–567
33. Deypir M, Sadreddini MH (2011) Eclats: an efficient sliding window based frequent pattern mining method for data streams. *Intell Data Anal* 15(4):571–587
34. Deypir M, Sadreddini MH, Hashemi S (2012) Towards a variable size sliding window model for frequent itemset mining over data streams. *Comput Ind Eng* 63(1):161–172
35. Elfeky MG, Aref WG, Elmagarmid AK (2005) Periodicity detection in time series databases. *IEEE Trans Knowl Data Eng* 17(7):875–887
36. Elgaml EM, Ibrahim DM, Sallam EA (2015) Improved fp-growth algorithm with multiple minimum supports using maximum constraints. *World Acad Sci Eng Technol Int J Comput Electr Autom Control Inf Eng* 9(5):1087–1094
37. Ezeife CI, Su Y (2002) Mining incremental association rules with generalized fp-tree. *Conference of the Canadian society for computational studies of intelligence*. Springer, Berlin, pp 147–160
38. Feldman R, Netzer O, Peretz A, Rosenfeld B (2015) Utilizing text mining on online medical forums to predict label change due to adverse drug reactions. In: *Proceedings of the 21th ACM SIGKDD international conference on knowledge discovery and data mining*, ACM, pp 1779–1788
39. Garofalakis M, Rastogi R, Shim K (2002) Mining sequential patterns with regular expression constraints. *IEEE Trans Knowl Data Eng* 14(3):530–552
40. Ge J, Xia Y (2016) Distributed sequential pattern mining in large scale uncertain databases. *Pacific-Asia conference on knowledge discovery and data mining*. Springer, Berlin, pp 17–29
41. Ge J, Xia Y, Wang J, Nadungodage CH, Prabhakar S (2016) Sequential pattern mining in databases with temporal uncertainty. *Knowl Inf Syst* 51:1–30
42. Giannella C, Han J, Pei J, Yan X, Yu PS (2003) Mining frequent patterns in data streams at multiple time granularities. *Next Gen Data Min* 212:191–212
43. Gu CK, Dong XL (2009) Efficient mining of local frequent periodic patterns in time series database. In: *2009 International conference on machine learning and cybernetics*, IEEE, vol 1, pp 183–186
44. Gupta A, Mittal A, Bhattacharya A (2011) Minimally infrequent itemset mining using pattern-growth paradigm and residual trees. In: *Proceedings of the 17th international conference on management of data*, computer society of India, p 13
45. Haglin DJ, Manning AM (2007) On minimal infrequent itemset mining. In: *DMIN*, pp 141–147
46. Han J, Fu Y (1995) Discovery of multiple-level association rules from large databases. *VLDB* 95:420–431
47. Han J, Gong W, Yin Y (1998) Mining segment-wise periodic patterns in time-related databases. In: *KDD*, pp 214–218
48. Han J, Dong G, Yin Y (1999) Efficient mining of partial periodic patterns in time series database. In: *Data engineering, 1999. Proceedings., 15th international conference on*, IEEE, pp 106–115
49. Han J, Pei J, Mortazavi-Asl B, Chen Q, Dayal U, Hsu MC (2000a) Freespan: frequent pattern-projected sequential pattern mining. In: *Proceedings of the sixth ACM SIGKDD international conference on Knowledge discovery and data mining*, ACM, pp 355–359
50. Han J, Pei J, Yin Y (2000b) Mining frequent patterns without candidate generation. *ACM Sigmod Rec ACM* 29:1–12
51. Han J, Pei J, Mortazavi-Asl B, Pinto H, Chen Q, Dayal U, Hsu M (2001) Prefixspan: Mining sequential patterns efficiently by prefix-projected pattern growth. In: *Proceedings of the 17th international conference on data engineering*, pp 215–224
52. Han J, Pei J, Kamber M (2011) *Data mining: concepts and techniques*. Elsevier, Oxford
53. Hemalatha CS, Vaidehi V, Lakshmi R (2015) Minimal infrequent pattern based approach for mining outliers in data streams. *Expert Syst Appl* 42(4):1998–2012
54. Hirate Y, Yamana H (2006) Sequential pattern mining with time intervals. *Pacific-Asia conference on knowledge discovery and data mining*. Springer, Berlin, pp 775–779
55. Hong TP, Lin CW, Wu YL (2008) Incrementally fast updated frequent pattern trees. *Expert Syst Appl* 34(4):2424–2435
56. Hoque N, Nath B, Bhattacharyya D (2013) An efficient approach on rare association rule mining. In: *Proceedings of seventh international conference on bio-inspired computing: theories and applications (BIC-TA 2012)*. Springer, Berlin, pp 193–203
57. Huan J, Wang W, Prins J (2003) Efficient mining of frequent subgraphs in the presence of isomorphism. In: *Data mining, 2003. ICDM 2003. Third IEEE international conference on*, IEEE, pp 549–552
58. Huang D, Koh YS, Dobbie G (2012) Rare pattern mining on data streams. In: *International conference on data warehousing and knowledge discovery*, Springer, pp 303–314
59. Huang DTJ, Koh YS, Dobbie G, Pears R (2014) Detecting changes in rare patterns from data streams. In: *Pacific-Asia conference on knowledge discovery and data mining*, Springer, pp 437–448
60. Huang KY, Chang CH (2004) Asynchronous periodic patterns mining in temporal databases. In: *Databases and applications*, pp 43–48
61. Huang KY, Chang CH (2005) Smca: a general model for mining asynchronous periodic patterns in temporal databases. *IEEE Trans Knowl Data Eng* 17(6):774–785
62. Huang Y, Zhang L, Zhang P (2008) A framework for mining sequential patterns from spatio-temporal event data sets. *IEEE Trans Knowl Data Eng* 20(4):433–448
63. Huang YF, Lin SY (2003) Mining sequential patterns using graph search techniques. In: *Computer software and applications conference, 2003. COMPSAC 2003. Proceedings. 27th annual international*, IEEE, pp 4–9
64. Hui L, Chen YC, Weng JTY, Lee SY (2016) Incremental mining of temporal patterns in interval-based database. *Knowl Inf Syst* 46(2):423–448
65. Huo W, Feng X, Zhang Z (2016) An efficient approach for incremental mining fuzzy frequent itemsets with fp-tree. *Int J Uncertain Fuzziness Knowl Based Syst* 24(03):367–386
66. Inokuchi A, Washio T, Motoda H (2000) An apriori-based algorithm for mining frequent substructures from graph data. In:

- European conference on principles of data mining and knowledge discovery, Springer, pp 13–23
67. Inokuchi A, Washio T, Motoda H (2003) Complete mining of frequent patterns from graphs: mining graph data. *Mach Learn* 50(3):321–354
  68. Ji Y, Ying H, Tran J, Dews P, Mansour A, Massanari RM (2013) A method for mining infrequent causal associations and its application in finding adverse drug reaction signal pairs. *IEEE Trans Knowl Data Eng* 25(4):721–733
  69. Jin R, Agrawal G (2005) An algorithm for in-core frequent itemset mining on streaming data. In: Fifth IEEE international conference on data mining (ICDM'05), IEEE, pp 8–pp
  70. Jin R, Agrawal G (2007) Frequent pattern mining in data streams. In: *Data streams*, Springer, pp 61–84
  71. Jindal R, Borah MD (2016) A novel approach for mining frequent patterns from incremental data. *Int J Data Min Model Manag* 8(3):244–264
  72. Kaneiwa K, Kudo Y (2011) A sequential pattern mining algorithm using rough set theory. *Int J Approx Reason* 52(6):881–893
  73. Kemmar A, Lebbah Y, Loudni S, Boizumault P, Charnois T (2016) Prefix-projection global constraint and top-k approach for sequential pattern mining. *Constraints* 22:1–42
  74. Kiran RU, Krishna Re P (2009) An improved multiple minimum support based approach to mine rare association rules. In: *Computational intelligence and data mining, 2009. CIDM'09. IEEE symposium on*, IEEE, pp 340–347
  75. Kiran RU, Reddy PK (2010) An efficient approach to mine rare association rules using maximum items support constraints. In: *British national conference on databases*, Springer, pp 84–95
  76. Kiran RU, Kitsuregawa M, Reddy PK (2016) Efficient discovery of periodic-frequent patterns in very large databases. *J Syst Softw* 112:110–121
  77. Koh YS, Ravana SD (2016) Unsupervised rare pattern mining: a survey. *ACM Trans Knowl Discov Data (TKDD)* 10(4):45
  78. Koperski K, Han J (1995) Discovery of spatial association rules in geographic information databases. In: *International symposium on spatial databases*, Springer, pp 47–66
  79. Kum HC, Pei J, Wang W, Duncan D (2003) Approxmap: Approximate mining of consensus sequential patterns. In: *SDM, SIAM*, pp 311–315
  80. Kuramochi M, Karypis G (2001) Frequent subgraph discovery. In: *Data mining, 2001. ICDM 2001. Proceedings IEEE international conference on*, IEEE, pp 313–320
  81. Kuramochi M, Karypis G (2005) Finding frequent patterns in a large sparse graph. *Data Min Knowl Discov* 11(3):243–271
  82. Lam HT, Mörchen F, Fradkin D, Calders T (2014) Mining compressing sequential patterns. *Stat Anal Data Min* 7(1):34–52
  83. Lavergne J, Benton R, Raghavan V, Hafez A (2013) Dyntarm: an in-memory data structure for targeted strong and rare association rule mining over time-varying domains. In: *Proceedings of the 2013 IEEE/WIC/ACM international joint conferences on web intelligence (WI) and intelligent agent technologies (IAT)-volume 01*, IEEE computer society, pp 298–306
  84. Lee G, Yun U, Ryu KH (2014) Sliding window based weighted maximal frequent pattern mining over data streams. *Expert Syst Appl* 41(2):694–708
  85. Lee YC, Hong TP, Lin WY (2005) Mining association rules with multiple minimum supports using maximum constraints. *Int J Approx Reason* 40(1):44–54
  86. Leleu M, Rigotti C, Boulicaut JF, Euvarard G (2003) Go-spade: mining sequential patterns over datasets with consecutive repetitions. In: *International workshop on machine learning and data mining in pattern recognition*, Springer, pp 293–306
  87. Leung CKS, Jiang F (2011) Frequent itemset mining of uncertain data streams using the damped window model. In: *Proceedings of the 2011 ACM symposium on applied computing*, ACM, pp 950–955
  88. Leung CKS, Khan QI (2006) Dstree: a tree structure for the mining of frequent sets from data streams. In: *Sixth international conference on data mining (ICDM'06)*, IEEE, pp 928–932
  89. Leung CKS, Khan QI, Li Z, Hoque T (2007) Cantree: a canonical-order tree for incremental frequent-pattern mining. *Knowl Inf Syst* 11(3):287–311
  90. Leung CKS, Cuzzocrea A, Jiang F (2013) Discovering frequent patterns from uncertain data streams with time-fading and landmark models. In: *Transactions on large-scale data-and knowledge-centered systems VIII*, Springer, pp 174–196
  91. Li HF, Lee SY (2009) Mining frequent itemsets over data streams using efficient window sliding techniques. *Expert Syst Appl* 36(2):1466–1477
  92. Li J, Zou Z, Gao H (2012) Mining frequent subgraphs over uncertain graph databases under probabilistic semantics. *VLDB J* 21(6):753–777
  93. Lin CH, Chiu DY, Wu YH, Chen AL (2005) Mining frequent itemsets from data streams with a time-sensitive sliding window. In: *SDM, SIAM*, pp 68–79
  94. Lin KW, Chung SH, Huang SS, Lin CC (2015) Fast mining frequent patterns with secondary memory. In: *Proceedings of the ASE bigdata & socialinformatics 2015*, ACM, p 43
  95. Lin MY, Lee SY (2003) Improving the efficiency of interactive sequential pattern mining by incremental pattern discovery. In: *System sciences, 2003. Proceedings of the 36th annual Hawaii international conference on*, IEEE, pp 8–pp
  96. Liu B, Hsu W, Ma Y (1999) Mining association rules with multiple minimum supports. In: *Proceedings of the fifth ACM SIGKDD international conference on knowledge discovery and data mining*, ACM, pp 337–341
  97. Liu G, Li J, Wong L, Hsu W (2006) Positive borders or negative borders: How to make lossless generator based representations concise. In: *SDM, SIAM*, pp 469–473
  98. Liu X, Mao GJ, Sun Y, Liu CN (2007) An algorithm to approximately mine frequent closed itemsets from data streams. *Dianzi Xuebao (Acta Electronica Sinica)* 35(5):900–905
  99. Liu X, Guan J, Hu P (2009) Mining frequent closed itemsets from a landmark window over online data streams. *Comput Math Appl* 57(6):927–936
  100. Ma S, Hellerstein JL (2001) Mining partially periodic event patterns with unknown periods. In: *Data engineering, 2001. Proceedings. 17th international conference on*, IEEE, pp 205–214
  101. Manku GS, Motwani R (2002) Approximate frequency counts over data streams. In: *Proceedings of the 28th international conference on very large data bases, VLDB endowment*, pp 346–357
  102. Marascu A, Masegla F (2006) Mining sequential patterns from data streams: a centroid approach. *J Intell Inf Syst* 27(3):291–307
  103. Masegla F, Cathala F, Poncelet P (1998) The psp approach for mining sequential patterns. In: *European symposium on principles of data mining and knowledge discovery*, Springer, pp 176–184
  104. Masegla F, Poncelet P, Teisseire M (2004) Pre-processing time constraints for efficiently mining generalized sequential patterns. In: *Temporal representation and reasoning, 2004. TIME 2004. Proceedings. 11th international symposium on*, IEEE, pp 87–95
  105. Metwally A, Agrawal D, El Abbadi A (2005) Efficient computation of frequent and top-k elements in data streams. In: *International conference on database theory*, Springer, pp 398–412
  106. Muzammal M, Raman R (2011) Mining sequential patterns from probabilistic databases. In: *Pacific-Asia conference on knowledge discovery and data mining*, Springer, pp 210–221
  107. Nguyen SN, Sun X, Orłowska ME (2005) Improvements of inc-span: Incremental mining of sequential patterns in large database.

- In: Pacific-Asia conference on knowledge discovery and data mining, Springer, pp 442–451
108. Nishi MA, Ahmed CF, Samiullah M, Jeong BS (2013) Effective periodic pattern mining in time series databases. *Expert Syst Appl* 40(8):3015–3027
  109. Nori F, Deypir M, Sadreddini MH (2013) A sliding window based algorithm for frequent closed itemset mining over data streams. *J Syst Softw* 86(3):615–623
  110. Pan F, Cong G, Tung AK, Yang J, Zaki MJ (2003) Carpenter: finding closed patterns in long biological datasets. In: Proceedings of the ninth ACM SIGKDD international conference on knowledge discovery and data mining. ACM, pp 637–642
  111. Pan F, Tung AK, Cong G, Xu X (2004) Cobbler: combining column and row enumeration for closed pattern discovery. In: Scientific and statistical database management, 2004. Proceedings. 16th international conference on, IEEE, pp 21–30
  112. Pant N, Kant S, Pant B, Sharma SK (2016) An efficient approach for mining sequential pattern. In: Proceedings of fifth international conference on soft computing for problem solving, Springer, pp 587–596
  113. Papapetrou O, Ioannou E, Skoutas D (2011) Efficient discovery of frequent subgraph patterns in uncertain graph databases. In: Proceedings of the 14th international conference on extending database technology, ACM, pp 355–366
  114. Pei J, Han J, Mortazavi-Asl B, Zhu H (2000) Mining access patterns efficiently from web logs. In: Pacific-Asia conference on knowledge discovery and data mining, Springer, pp 396–407
  115. Pei J, Han J, Lu H, Nishio S, Tang S, Yang D (2001) H-mine: hyper-structure mining of frequent patterns in large databases. In: Data mining, 2001. ICDM 2001, proceedings IEEE international conference on, IEEE, pp 441–448
  116. Pei J, Han J, Wang W (2002) Mining sequential patterns with constraints in large databases. In: Proceedings of the eleventh international conference on Information and knowledge management, ACM, pp 18–25
  117. Pei J, Han J, Wang W (2007) Constraint-based sequential pattern mining: the pattern-growth methods. *J Intell Inf Syst* 28(2):133–160
  118. Peng WC, Liao ZX (2009) Mining sequential patterns across multiple sequence databases. *Data Knowl Eng* 68(10):1014–1033
  119. Pijls W, Bioch JC (2000) Mining frequent itemsets in memory-resident databases. ERIM report series reference no ERS-2000-53-LIS
  120. Pillai J, Vyas O (2011) High utility rare item set mining (huri): An approach for extracting high utility rare item sets. *i-Manag J Future Eng Technol* 7(1):25
  121. Pinto H, Han J, Pei J, Wang K, Chen Q, Dayal U (2001) Multi-dimensional sequential pattern mining. In: Proceedings of the tenth international conference on Information and knowledge management, ACM, pp 81–88
  122. Plantevit M, Choong YW, Laurent A, Laurent D, Teisseire M (2005) M2sp: mining sequential patterns among several dimensions. In: European conference on principles of data mining and knowledge discovery, Springer, pp 205–216
  123. Rahman A, Ezeife CI, Aggarwal AK (2010) Wifi miner: an online apriori-infrequent based wireless intrusion system. In: Knowledge discovery from sensor data, Springer, pp 76–93
  124. Rahman A, Xu Y, Radke K, Foo E (2016) Finding anomalies in scada logs using rare sequential pattern mining. In: International conference on network and system security, Springer, pp 499–506
  125. Raïssi C, Poncelet P, Teisseire M (2007) Towards a new approach for mining frequent itemsets on data stream. *J Intell Inf Syst* 28(1):23–36
  126. Rasheed F, Alshalalfa M, Alhadj R (2011) Efficient periodicity mining in time series databases using suffix trees. *IEEE Trans Knowl Data Eng* 23(1):79–94
  127. Sadhasivam KS, Angamuthu T (2011) Mining rare itemset with automated support thresholds. *J Comput Sci* 7(3):394
  128. Saha B, Lazarescu M, Venkatesh S (2007) Infrequent item mining in multiple data streams. In: Seventh IEEE international conference on data mining workshops (ICDMW 2007), IEEE, pp 569–574
  129. Selvi CK, Tamilarasi A (2009) Mining association rules with dynamic and collective support thresholds. *Int J Eng Technol* 1(3):236–240
  130. Seno M, Karypis G (2002) Slpminer: an algorithm for finding frequent sequential patterns using length-decreasing support constraint. In: Data mining, 2002. ICDM 2002. Proceedings. 2002 IEEE international conference on, IEEE, pp 418–425
  131. Sheng C, Hsu W, Lee ML (2006) Mining dense periodic patterns in time series data. In: 22nd international conference on data engineering (ICDE'06), IEEE, pp 115–115
  132. Srikant R, Agrawal R (1995) Mining generalized association rules. IBM Research Division, Zurich
  133. Srikant R, Agrawal R (1996) Mining sequential patterns: generalizations and performance improvements. In: International conference on extending database technology, Springer, pp 1–17
  134. Sunitha G, Reddy ARM (2016) Wrsp-miner algorithm for mining weighted sequential patterns from spatio-temporal databases. In: Proceedings of the second international conference on computer and communication technologies, Springer, pp 309–317
  135. Szathmary L, Napoli A, Valtchev P (2007) Towards rare itemset mining. In: Tools with artificial intelligence, 2007. ICTAI 2007. 19th IEEE international conference on, IEEE, vol 1, pp 305–312
  136. Talukder N, Zaki M (2016) A distributed approach for graph mining in massive networks. *Data Min Knowl Discov* 30:1–29
  137. Tanbeer SK, Ahmed CF, Jeong BS, Lee YK (2008a) Cp-tree: a tree structure for single-pass frequent pattern mining. In: Pacific-Asia conference on knowledge discovery and data mining, Springer, pp 1022–1027
  138. Tanbeer SK, Ahmed CF, Jeong BS, Lee YK (2008b) Efficient frequent pattern mining over data streams. In: Proceedings of the 17th ACM conference on Information and knowledge management, ACM, pp 1447–1448
  139. Tanbeer SK, Ahmed CF, Jeong BS (2009a) Parallel and distributed algorithms for frequent pattern mining in large databases. *IETE Tech Rev* 26(1):55–66
  140. Tanbeer SK, Ahmed CF, Jeong BS, Lee YK (2009b) Efficient single-pass frequent pattern mining using a prefix-tree. *Inf Sci* 179(5):559–583
  141. Tanbeer SK, Ahmed CF, Jeong BS, Lee YK (2009c) Sliding window-based frequent pattern mining over data streams. *Inf Sci* 179(22):3843–3865
  142. Tao F, Murtagh F, Farid M (2003) Weighted association rule mining using weighted support and significance framework. In: Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining, ACM, pp 661–666
  143. Teng WG, Chen MS, Yu PS (2003) A regression-based temporal pattern mining scheme for data streams. In: Proceedings of the 29th international conference on Very large data bases-volume 29, VLDB endowment, pp 93–104
  144. Troiano L, Scibelli G, Birtolo C (2009) A fast algorithm for mining rare itemsets. In: 2009 ninth international conference on intelligent systems design and applications, IEEE, pp 1149–1155
  145. Tsai PS (2009) Mining frequent itemsets in data streams using the weighted sliding window model. *Expert Syst Appl* 36(9):11617–11625
  146. Tsang S, Koh YS, Dobbie G (2011) Rp-tree: rare pattern tree mining. In: Data warehousing and knowledge discovery, Springer, pp 277–288
  147. Tzvetkov P, Yan X, Han J (2005) Tsp: mining top-k closed sequential patterns. *Knowl Inf Syst* 7(4):438–457

148. Van Wijk JJ, Van Selow ER (1999) Cluster and calendar based visualization of time series data. In: Information visualization, 1999.(Info Vis' 99) proceedings. 1999 IEEE symposium on, IEEE, pp 4–9
149. Vijayalakshmi R, Nadarajan R, Roddick JF, Thilaga M, Nirmala P (2011) Fp-graphminer—a fast frequent pattern mining algorithm for network graphs. *J Gr Algorithms Appl* 15(6):753–776
150. Wang C, Wang W, Pei J, Zhu Y, Shi B (2004) Scalable mining of large disk-based graph databases. In: Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining, ACM, pp 316–325
151. Wang J, Zeng Y (2011) Dswfp: Efficient mining of weighted frequent pattern over data streams. In: Fuzzy systems and knowledge discovery (FSKD), 2011 eighth international conference on, IEEE, vol 2, pp 942–946
152. Wang JZ, Huang JL (2016) Incremental mining of high utility sequential patterns in incremental databases. In: Proceedings of the 25th ACM international on conference on information and knowledge management, ACM, pp 2341–2346
153. Wang K, He Y, Han J (2003) Pushing support constraints into association rules mining. *IEEE Trans Knowl Data Eng* 15(3):642–658
154. Wang W, Wang C, Zhu Y, Shi B, Pei J, Yan X, Han J (2005) Graphminer: a structural pattern-mining system for large disk-based graph databases and its applications. In: Proceedings of the 2005 ACM SIGMOD international conference on Management of data, ACM, pp 879–881
155. Weiss SM (2004) A framework for discovering co-location patterns in data sets with extended spatial objects
156. Wong RCW, Fu AWC (2006) Mining top-k frequent itemsets from data streams. *Data Min Knowl Discov* 13(2):193–217
157. Xu B, Yi T, Wu F, Chen Z (2002) An incremental updating algorithm for mining association rules. *J Electron (China)* 19(4):403–407
158. Yan X, Han J (2002) gspan: graph-based substructure pattern mining. In: Data mining, 2002. ICDM 2003. Proceedings. 2002 IEEE international conference on, IEEE, pp 721–724
159. Yan X, Han J (2003) Closegraph: mining closed frequent graph patterns. In: Proceedings of the ninth ACM SIGKDD international conference on knowledge discovery and data mining, ACM, pp 286–295
160. Yan X, Cheng H, Han J, Yu PS (2008) Mining significant graph patterns by leap search. In: Proceedings of the 2008 ACM SIGMOD international conference on management of data, ACM, pp 433–444
161. Yang J, Wang W, Yu PS, Han J (2002) Mining long sequential patterns in a noisy environment. In: Proceedings of the 2002 ACM SIGMOD international conference on management of data, ACM, pp 406–417
162. Yang J, Wang W, Yu PS (2003) Mining asynchronous periodic patterns in time series data. *IEEE Trans Knowl Data Eng* 15(3):613–628
163. Yang Z, Kitsuregawa M (2005) Lapin-spam: an improved algorithm for mining sequential pattern. In: 21st International conference on data engineering workshops (ICDEW'05), IEEE, pp 1222–1222
164. Yu CC, Chen YL (2005) Mining sequential patterns from multidimensional sequence data. *IEEE Trans Knowl Data Eng* 17(1):136–140
165. Yu JX, Chong Z, Lu H, Zhou A (2004) False positive or false negative: mining frequent itemsets from high speed transactional data streams. In: Proceedings of the thirtieth international conference on very large data bases-volume 30, VLDB endowment, pp 204–215
166. Yun H, Ha D, Hwang B, Ryu KH (2003) Mining association rules on significant rare data using relative support. *J Syst Softw* 67(3):181–191
167. Yun U, Lee G (2016a) Incremental mining of weighted maximal frequent itemsets from dynamic databases. *Expert Syst Appl* 54:304–327
168. Yun U, Lee G (2016b) Sliding window based weighted erasable stream pattern mining for stream data applications. *Future Gen Comput Syst* 59:1–20
169. Yun U, Lee G, Ryu KH (2014) Mining maximal frequent patterns by considering weight conditions over data streams. *Knowl Based Syst* 55:49–65
170. Zhou D, Karthikeyan A, Wang K, Cao N, He J (2016) Discovering rare categories from graph streams. *Data Min Knowl Discov* 31:1–24
171. Zhu F, Yan X, Han J, Philip SY (2007a) gprune: a constraint pushing framework for graph pattern mining. In: Pacific-Asia conference on knowledge discovery and data mining, Springer, pp 388–400
172. Zhu F, Yan X, Han J, Philip SY, Cheng H (2007b) Mining colossal frequent patterns by core pattern fusion. In: 2007 IEEE 23rd international conference on data engineering, IEEE, pp 706–715
173. Zhu J, Wang K, Wu Y, Hu Z, Wang H (2016) Mining user-aware rare sequential topic patterns in document streams. *IEEE Trans Knowl Data Eng* 28(7):1790–1804
174. Zou Z, Gao H, Li J (2010a) Discovering frequent subgraphs over uncertain graph databases under probabilistic semantics. In: Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining, ACM, pp 633–642
175. Zou Z, Li J, Gao H, Zhang S (2010b) Mining frequent subgraph patterns from uncertain graph data. *IEEE Trans Knowl Data Eng* 22(9):1203–1218

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.