Journal of
Data Acquisition and Processing
ISSN: 1004-9037
https://sjcjycl.cn/
DOI: 10.5281/zenodo.777597

STUDY OF FREQUENT PATTERN MINING TECHNIQUES, APPLICATIONS AND PERFORMANCE

Arpit Solanki and Dr. Rajeev G. Vishwakarma
Department of Computer Science & Engineering,
Dr. A. P. J. Abdul Kalam University, Indore (M.P.) - 452010, India
Corresponding Author Email : arpit.solanki29@gmail.com

Abstract: Data mining is a classical research area. That is utilized for analyzing data for prepare a model which will able to deal with the classification problems, categorization problems and rule building techniques. The rule building techniques are basically utilized to understand or map the attribute relationships for making smart decisions. Among different rule mining techniques association rule mining or frequent pattern mining technique is one of the popular techniques. In this presented work, the frequent pattern mining algorithm is the key area of investigation.Therefore first a review has been carried out for identifying the recent applications of frequent pattern mining algorithms. Next three popular frequent pattern mining algorithms are considered for experimental study. The experiments on publically available UCI based datasets has been carried out and the performance of the algorithms has been measured and compared with each other. According to the performance measurement we found that the éclat is efficient and effective algorithm for frequent pattern mining. But the apriori algorithm is most promising algorithm among the compared algorithms. Therefore, based on experiments the key problem has been identified and future extension of the work has been proposed.
Keywords: Frequent pattern mining, Association rule mining, Data mining, Rule mining techniques, experimental comparison.

INTRODUCTION

The data mining is one of the traditional research domains which are used for data analysis and pattern recovery. The automated algorithms and application oriented data has been used to explore and recover the essential insights from the data. In this context, there are a number of different algorithms are available for performing classification, clustering, prediction and relationship building. Among the different type of algorithms the relationship building and understanding of data is one of the essential tasks. Therefore in this proposed work the main aim is to explore and understand the technique of frequent pattern mining.

Frequent pattern mining is the process of identifying patterns within a dataset that occur frequently. This is done by analyzing datasets to find items that appear together.This technique is useful for discovering that information which is not visible in data. Additionally, it can find association and correlation among data items or attributes. But there are some limitations are also available such as it can generate a large number of patterns with high dimensionality. Additionally, the number of patterns can be very large, making it difficult to interpret the results.in addition, there are a number of applications where the frequent pattern mining can be employed, some essential of them are given as:

- Market Basket Analysis:In this is application algorithm analyzing customer purchasing patterns to identify items that are frequently bought together. This information can be used to optimize product placement, create marketing campaigns, and make business decisions.

- Recommender Systems:It can be used to identify patterns according to user behavior and preferences to make appropriate product suggestions.
- Network Intrusion Detection: Network administrators are utilizing these algorithms to detect network activity to identify a security threat.
- Medical Analysis:This algorithm can be used to identify patterns in medical data that may indicate a particular disease.
- Text Mining:It can be used to identify patterns in text, such as keywords or phrases that appear frequently in a document.
- Web usage mining:It can be used to analyze patterns of user behavior on a website, such as which pages are visited most frequently or links clicked.

This section discusses the basics of the frequent pattern mining and the applications. The next section includes the related study which is recently contributed for improving the frequent pattern analysis algorithm.

RELATED STUDY

This section offers the background study related to the frequent pattern mining and association rule mining.

Essential Keywords

Table 1 List of keywords

Abbreviation	Full form
ARM	Association Rule Mining
FPM	Frequent Patterns Mining
SPM	Sequential Pattern Mining
NetNCSP	Nettreefor Non-Overlapping Closed Sequential Pattern
GPM	Graph Pattern Mining
HUOPM	High Utility Occupancy Pattern Mining
FU-tree	Frequency-Utility Tree
PSPM	Parallel Sequential Pattern Mining
SPP	Safe Pattern Pruning
FSG	Frequent Sub Graph Discovery
DNA	Deoxyribonucleic Acid
TOU	Time Of Use
PDR	Peak Demand Reduction

Recent Review

High utility mining is an emerging data science task, aims to extract knowledge based on a objective. The utility shows its benefit that can be calculated based on user priority and domain understanding. Mining itemsets in big data have received attention based on the Apache Hadoop and Spark.S. Kumar et al [1]give an overview of the distinct approaches to pattern mining in Big Data. They investigate the problem involved with pattern mining and associatedtechniques such as Hadoop, Spark, parallel and distributed processing. Then, examine developments in parallel, distributed, and scalable pattern mining, analyze them in the big dataperspective and identify difficulties. They study four varietiesof itemsets mining, i.e., parallel frequent itemsets mining, high utility mining, sequential patternsmining and frequent itemset mining. They conclude with openissues and opportunity. It also provides direction for enhancement.

F. Min et al [2] proposes a frequent pattern discovery algorithm for a new type of pattern by dividing the alphabet into strong, medium, and weak parts. It is called a tri-pattern.It is general and flexible, therefore more interesting. Experiments were undertaken on data in various fields to reveal the universality of pattern. These include protein sequence, petroleum production time series, and forged Chinese text. The results show that tri-patterns are meaningful than existing types of patterns. Itenriches the semantics of SPM and applications.

The hidden patterns of the frequent itemsets become more time consuming when the amount of data increases. Moreover, memory is needed in mining the hidden patterns due to a heavy computation. Therefore, an efficient algorithm is required to mine frequent itemsets within a shorter run time and with less memory while the volume of data increases. C. H. Chee et al [3] presents a comparison of different algorithms for Frequent Pattern Mining so that a more efficient algorithm can be developed.

SPM neglects the repetition in sequence. To solve this problem, gap constraint SPM was proposed by Y. Wu et al [4] and avoids finding too many useless patterns. Non-overlapping SPM, means that any two occurrences cannot use the same sequence letter in same position. Non-overlapping SPM can balance efficiency and completeness. The existing FPM methods contain redundant patterns. To reduce them and improve performance, author adopts closed pattern mining and proposes a complete algorithm, named NetNCSP. It is equipped with two steps, support calculation and closeness determination. Backtracking strategy is used to calculate non-overlapping support on the corresponding Nettree, which reduces time complexity. They also propose three kinds of pruning, inheriting, predicting, and determining. The pruning strategies are able to find the redundant patterns and can predict the frequency and closeness before the generation of the candidates. Results show that NetNCSP is not only efficient but also discover closed patterns. It mines the closed patterns in SARS-CoV-2 and SARS viruses.

V. Dias et al [5] propose Fractal, a high performance and high productivity system for supporting distributed GPM. Fractal employs a dynamic load-balancing based on a hierarchical and locality-aware work stealing mechanism, to adapt workload characteristics. Additionally, it enumerates sub-graphs by combining a depth-first strategy with scratch processing paradigm to avoid storing large amounts of intermediate state andimproves memory efficiency. For programmer productivity, it presents an intuitive, expressive and modular API. Fractal-based implementations outperform both existing solutions on many problems to sub-graph querying.

To extract high quality patterns, W. Gan et al [6] extends the occupancy measure to assess the utility of patterns. They propose an algorithm named HUOPM. It considers user preferences like frequency, utility, and occupancy. A FU-tree and two data structures, called the utility occupancy list and FU-table, are designed for pruning. This method can discover the complete set of high quality patterns without candidate generation. Experiments have been conducted on several datasets. Results show that the patterns are intelligible, reasonable and acceptable, and with its pruning it outperforms, in terms of runtime and search space.

Data mining algorithms have problems like memory cost, processing speed, and space. The FPM and ARM, suffers from these challenges. W. Gan et al [7], survey current status of PSPM is investigated, including categorization and parallel SPM. They review parallel SPM in detail, including partition, Apriori, pattern growth, and hybrid algorithms, and provide characteristics, advantages, disadvantages and summarization. Topics, including parallel quantitative / weighted / utility SPM, uncertain data and stream data, hardware acceleration, are reviewed.

R. Bunker et al [8] apply a recent supervised sequential pattern mining algorithm called SPP to 490 labeled event sequences representing passages of play from one rugby team's matches inJapan. They obtain patterns that are the discriminative between scoring and non-scoring outcomes from both the team's and opposition teams' perspectives, and compare with the most frequent patterns when applied to subsets of the dataset. From obtained results, line breaks, successful line-outs, regained kicks in play, repeated phase-breakdown, and failed plays by the opposition team were found to be the patterns that discriminated most between the team scoring and not scoring. It was also found that, because of the supervised nature and pruning mechanisms of SPP, compared to the patterns obtained by the unsupervised methods, were more sophisticated in terms of containing a greater variety of events, and interpreted, the SPP-obtained patterns would be useful for coaches and performance analysts.

T. A. D. Lael et al [9] analyze consumer purchasing patterns for motorcycle parts using FP-Growth algorithm on sales transaction. The aim is to obtain information for companies in planning marketing strategies and increasing sales. The data used are motorcycle parts sales data from motorcycle parts stores. The data is then processed using the FP-Growth algorithm to find significant patterns. The results show that the FP-Growth algorithm can be used to identify substantial consumer purchasing patterns. Some patterns found a combination of often purchased products, active purchase time, and product category. Using FP-Growth algorithm can assist companies in understanding consumer purchasing patterns to improve the marketing strategies and increase sales. The novelty lies in using data mining methods and FP-Growth on motorcycle parts sales transaction data. This is also identifying purchasing patterns, such as product combinations purchased together and product categories.

The objective of C. R. Wijesinghe et al [10]is to identify the frequent workflow patterns in a corpus of Galaxy bioinformatics workflows. FSG algorithm is used in analyzing the workflows. Seventy-one reusable workflow patterns identified with a 5% minimum support. Future plan is to annotate the identified frequent patterns and encode the patterns in the workflow systems with the objective of improving the usability by a high-level interface to the user.

DNA main function is information storage. The advancement of sequencing technology had caused DNA sequence data to grow at an explosive rate, which has pushed the study in the big data. Moreover, Machine Learning (ML) is a technique for analyzing largescale data and learns to gain knowledge. A. Yang et al [11] introduces development process of sequencing technology, expounds on the DNA sequence data structure and similarity. They analyze process of data mining, summary ML algorithms, and put forward the challenges ofML algorithms in the mining of biological sequence. Then, review four applications of ML in DNA sequence data: sequence alignment, classification, clustering, and pattern mining. They analyze corresponding biological application background and significance, and summarized development and problems. Finally, summarize the content of the review and look into the future directions.

F. Wang et al [12], an ARM based analysis framework is built to explore the impacts of characteristics on PDR under TOU price. First, a customer PDR characterizing model is proposed, where difference-in-difference model is adopted to quantify the effect and probability distribution fitting method is used to characterize the feature. Then anARM analysis using Apriori algorithm is presented to explore the impactscovering four categories: dwelling characteristics, socio-demographic, appliances and heating, and attitudes towards energy. Finally, results based on 2993 records containing smart metering data illustrate that PDR level cannot be obtained based on ownership and usage. Socio-demographic information Internet connection and house insulation contribute to the increase of PDR level. The percentage of renewable generation also shows a relationship. The framework improves the benefits of TOU programs and guide policy makers to design efficient energy saving policies.

Bilharzia or schistosomiasis is one of the most fatal and factitious disease becomes a reason of deaths. Prediction and factors identification that causes of disease in early stage, may escort to treatment. Data mining techniques are used to assist medical professionals in diseases' classification. Y. Ali et al [13] investigates the recovery and death factors which contributes to schistosomiasis disease dataset, collected from Hubei, China. A learning method, ARM (Apriori) is used to spot factors. Different tools were used for analysis and model evaluation with minimum support and minimum confidence indicated higher than 90% to generate rules. In addition, attributes indicating recovery and death of individuals were identified. Strong associations factors; BMI, viability, nourishment, extent to ascites etc. determined and classified. Results generated by ARM may useful for professionals in precise treatment decision.

According to M. H. Santoso et al [14]data mining can be defined as a technique for findingpatterns in large amounts ofdata for decision support. One of the commonly used ARM methodsis the Apriori. The Association Rule and the AprioriAlgorithm are two prominent algorithms for finding a numberof frequently occurring sets of items from transaction data. The calculation is done to determine the minimum valueof support and confidence that will produce theassociation rule. The rule is used to produce thepercentage of purchasing activity for an itemset within aperiod of time using RapidMiner. By searching for patterns using this apriori algorithm, it is hoped that the resulting information canimprove further sales strategies.

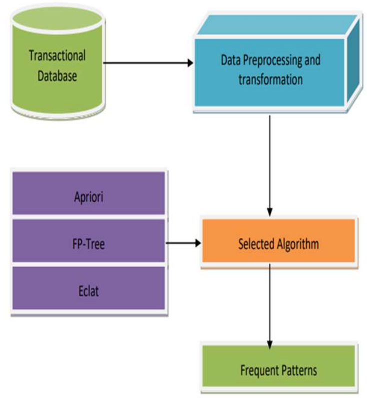


Figure 1: Proposed Experimental Model

According to S. Das et al [15]in 2011, 4,432 pedestrians were killed, and 69,000 pedestrians were injured in vehicle-pedestrian crashes in Louisiana,United States. Vehicle crashes have become concern because ofthe high percentage of fatalities. In 2012,pedestrians were accounted for 17%. Alcohol was involved in nearly 44%. Authors utilized 'Apriori' algorithm to discover patterns. They aim to discover vehicle-pedestrian crashpatterns using eight years of data (2004–2011).The results indicated that roadway lighting at night helped inalleviating pedestrian crash. A few groups ofinterest were identified: male pedestrians' greaterpropensity towards severe and fatal crashes, younger femaledrivers (15–24) being more crash-prone,vulnerable impaired pedestrians even on roadways with lightingat night, middle-aged male pedestrians (35–54) being inclinedtowards crash occurrence, and dominance of single vehiclecrashes. Based on the patterns, theyrecommend severalmeasures to alleviate the safety. The findings will help traffic safetyprofessionals in understanding patterns andcountermeasures to raise awareness and improvements for the decrease of pedestrian crashes.

PROPOSED WORK

The proposed work is motivated to study the performance influence of frequent pattern mining algorithms. In this context, an experimental model has been demonstrated in figure 1. It demonstrates the proposed experimental model for comparative performance study among the classical frequent pattern mining algorithms. In this context, first component of the model has established as the transactional database. The transactional databases are the growing database with the time and contain the records indicating an event or physical or virtual transactions. The database has been taken into consideration for extracting the meaningful patterns and insights. Therefore, it is required to employ the frequent pattern analysis we need to prepare the data for utilizing with the frequent pattern analysis technique. In this context, the data is needed to be transformed and preprocess.

The aim of preprocessing technique is to enhance the data quality and improve the learning efficiency and quality of service. But in frequent pattern mining techniques we need special concentration on preprocessing of transactional databases. In this context, the aim is to identify the Itemsets and transaction sets based on the dataset. Itemsets are the set of unique symbols which are used to craft the transactions. These transaction set are the group of items set for describing the sessions of a user. Therefore, the dataset is preprocessed and transformed into transaction sets and utilized with the frequent pattern mining algorithm. In this context, the apriori, FP-tree and éclat algorithm is considered. The brief overview of the frequent pattern mining algorithm is given as:

Apriori algorithm

Apriori algorithm is easy and very simple, to mine frequent itemsets in a transactional database. The algorithm makes searches in database to find frequent itemsets where k itemsets are used to generate k+1- itemsets. Each k-itemset must be greater than or equal to minimum support threshold to be frequency. Otherwise, it is called candidate itemsets. In the first, the algorithm scan database to find frequency of 1-itemsets that contains only one item by counting each item in database. The frequency of 1-itemsets is used to find the itemsets in 2- itemsets which in turn is used to find 3-itemsets and so on until there are not any more k-itemsets. If an itemset is not frequent, any large subset from it is also non-frequent; this condition prune from search space in database.

Table 1 Apriori Algorithm

Algorithm: Apriori (T, ϵ)

Process:
1. $L1 \leftarrow \{large\ 1 - itemsets\}$
2. $k \leftarrow 2$
3. $while\ L_{k-1} \neq \emptyset$
    a. $C_K \leftarrow \{c = a \cup \{b\} \mid a \in L_{k-1} \cap b \notin a, \{s \subseteq c \mid \lvert s \rvert = k - 1\} \subseteq L_{k-1}\}$
    b. $for\ transactions\ t\ \in T$
        i. $D_t \leftarrow \{c \in C_k \mid c \subseteq t\}$
        ii. $for\ candidates\ c \in D_t$
            1. $count[c] \leftarrow count\ [c] + 1$
        iii. End for
        iv. $L_k \leftarrow \{c \in C_k \mid count\ [c] \geq e\}$
    c. End for
4. $k = k + 1$
5. End while
6. $return\ \coprod_k L_k$

FP-Tree algorithm: Frequent Pattern Tree is a tree-like structure that is made with the initial itemsets of the database. The purpose of the FP tree is to mine the most frequent pattern. Each node of the FP tree represents an item of the itemset. The root node represents null while the lower nodes represent the itemsets. The association of the nodes with the lower nodes that is the itemsets with the other itemsets is maintained while forming the tree. The frequent pattern growth method lets us find the frequent pattern without candidate generation. Let us see the steps followed to mine the frequent pattern using frequent pattern growth algorithm:

The first step is to scan the database to find the occurrences of the itemsets in the database. This step is the same as the first step of Apriori. The count of 1-itemsets in the database is called support count or frequency of 1-itemset.

The second step is to construct the FP tree. For this, create the root of the tree. The root is represented by null.

The next step is to scan the database again and examine the transactions. Examine the first transaction and find out the itemset in it. The itemset with the max count is taken at the top, the next itemset with lower count and so on. It means that the branch of the tree is constructed with transaction itemsets in descending order of count.

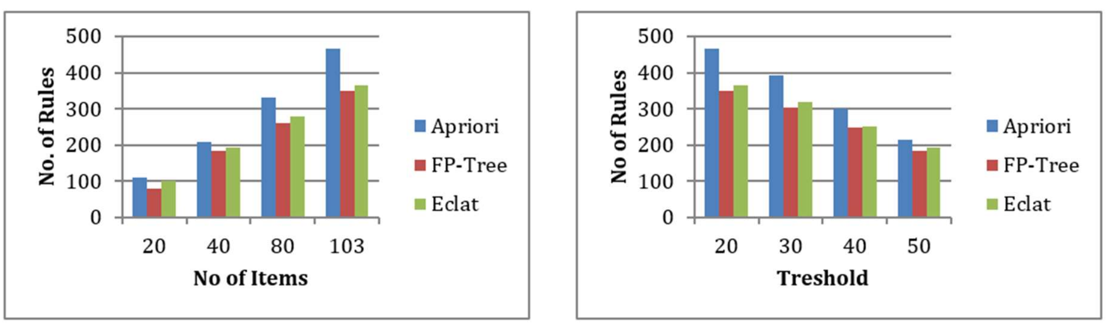
The next transaction in the database is examined. The itemsets are ordered in descending order of count. If any itemset of this transaction is already present in another branch (for example in the 1st transaction), then this transaction branch would share a common prefix to the root. This means that the common itemset is linked to the new node of another itemset in this transaction. Also, the count of the itemset is incremented as it occurs in the transactions. Both the common node and new node count is increased by 1 as they are created and linked according to transactions.

The next step is to mine the created FP Tree. For this, the lowest node is examined first along with the links of the lowest nodes. The lowest node represents the frequency pattern length 1. From this, traverse the path in the FP Tree. This path or paths are called a conditional pattern base. Conditional pattern base is a sub-database consisting of prefix paths in the FP tree occurring with the lowest node (suffix).

Construct a Conditional FP Tree, which is formed by a count of itemsets in the path. The itemsets meeting the threshold support are considered in the Conditional FP Tree.

Frequent Patterns are generated from the Conditional FP Tree [66].

Eclat Algorithm: The ECLAT algorithm stands for Equivalence Class Clustering and bottom-up Lattice Traversal. It is a popular method of Association Rule mining. It is a more efficient and scalable version of the Apriori algorithm. The Apriori algorithm works in a horizontal sense imitating the Breadth-First Search, the ECLAT algorithm works in a vertical manner like the Depth-First Search. This approach of the algorithm makes it a faster than the Apriori algorithm.The basic idea is to use Transaction Id intersections to compute the support of a candidate and avoiding the generation of subsets which are not in the prefix tree. In the first call of the function, all single items are used. Then the function is called recursively and in each recursive call, each item-tidset pair is verified and combined with other pairs.



(A) (B)

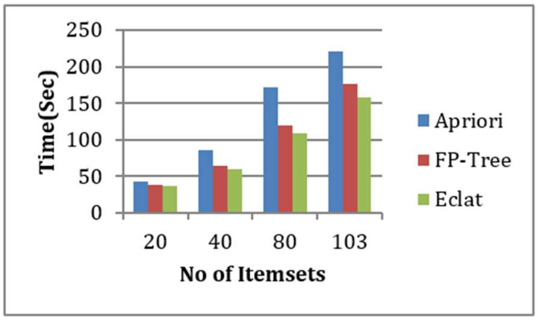
Figure 2: Number of generated frequent pattern rules by variations of (A) Number of itemsets (B) Filtering Threshold

This process is continued until no candidate item-tidset pairs can be combined. Advantages over Apriori algorithm:

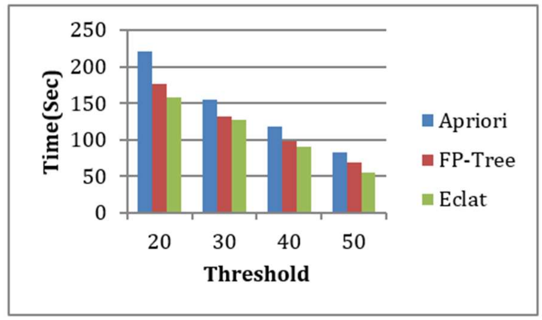
Memory Requirements: Since the ECLAT algorithm uses a Depth-First Search approach, it uses less memory than Apriori algorithm.

Speed: The ECLAT algorithm is typically faster than the Apriori algorithm.

Number of Computations: The ECLAT algorithm does not involve the repeated scanning of the data to compute the individual support values.



(A)



(B)

Figure 3: shows the training time for frequent rule mining techniques by variations of (A) Number of itemsets (B) Filtering Threshold

Next,a provision has been implemented to select an appropriate frequent pattern mining algorithm among three implemented algorithms. Additionally, using different size of datasets the experiments have been carried out and performance has been evaluated. Different performance parameters have been used to evaluate the models in terms of time and memory cost. The next section describes the obtained performance of the frequent pattern algorithms.

RESULTS ANALYSIS

The aim of the work is to investigate the performance of frequent rule mining algorithms. In this context, an experimental model has been implemented and the publically available frequent pattern datasets are utilized from UCI repository. Additionally the performance in terms of number of rules generated and time utilization to generate the rules has been evaluated. In this context, first the number of items is increasing for mining the rules additionally the number of rules and training time is recorded. Additionally in second scenario the support threshold has been increased and the number of rules and training time has been recorded.

Figure 2 demonstrate the performance of the different frequent pattern mining algorithms in terms of generated number of frequent pattern rules. Therefore, first the number of itemsets is varied and the number of rules has been generated. Figure 2(A) shows the number of rule generated according to the different number of itemsets. The X axis shows the number of datasets used for experiments and Y axis shows the number of rules generated according to the number of items. On the other hand the number of rule generated by increasing amount of filtering threshold or support values in terms of percentage (%). Figure 2(B) shows the number of rules generated by increasing amount of support. The X axis shows the threshold values and Y axis shows the number of rule generated. According to the results we found that the number of rules is increasing with the amount of items involved in experimental data. Additionally, when the support threshold has been increased then the number of generated rules is reducing in the similar ratio.

Next,the training time is also measured, which shows computational cost of the algorithm.It is the amount of time required to mine the frequent patterns. The time can be measured using:

$$time\ consumed = End\ time - Start\ Time$$

The time consumed is measured in terms of seconds (sec). Figure 3 shows the time consumed by different frequent pattern mining algorithms. Figure 3(A) includes the time consumed for processing the data with the increasing number of items. The X axis of diagram shows the size of itemsets used and Y axis shows the time utilized for extraction of frequent patterns. Similarly, in figure 3(B) shows the training time by varying the filtering or support threshold. In this diagram the Y axis contains the time taken and X axis shows the threshold. According to the results the training time is increasing with the amount itemsets and reducing with the increasing thresholds.

CONCLUSION & FUTURE WORK

In classical data mining for establishing relationship among frequent events association rule mining or frequent pattern mining techniques has been employed. The aim of this technique is to recover all the possible frequent events in the given database. Mostly these techniques are utilized with a transactional database. In order to calculate the frequent patterns a number of algorithms are exist but there are three key algorithms are available namely Apriori, FP-Tree and Eclat. In this presented work the aim is to explore the utilization or applications of frequent pattern mining algorithms. Therefore, review of recent applications of association rule mining and frequent pattern mining algorithms in real worlds has been conducted. Next, for studying the different frequent pattern algorithms the three popular algorithms has been considered namely Apriori, FP-Tree and Eclat.

Therefore, an experimental model has been configured and reported to conduct experiments on the different algorithms. Using the experiments with the dataset available in UCI repository based on the experiments results are summarized as:

1. The number of itemsets at influencing the performance in terms of number of rules as well as training time.
2. The increasing filtering threshold or support values can reduce the amount of time consumption and number of rules
3. The increasing support values can also responsible for information loss and negative impact on application's quality of service.
4. Performance of éclat algorithm is efficient as compared to FP-Tree and apriori algorithm in terms of both training time and number of rules

Based on collected and studied review and conducted reviews we found that the balance between information losses using the support is a major issue. Additionally, the long running time is also a key issue which is controlled by support value as requirement. Therefore we need a balance between efficiency and quality of service requirements. In near future we proposed a unique solution which is able to deal with the problem statement, and minimize the information loss which negatively impact of frequent pattern mining application.

REFERENCES

- S. Kumar, K. K. Mohbey, "A review on big data based parallel and distributed approaches of pattern mining", Journal of King Saud University – Computer and Information Sciences 34 (2022) 1639–1662

- F. Min, Z. H. Zhang, W. J. Zhai, R. P. Shen, “Frequent pattern discovery with tri-partition alphabets”, Information Sciences 000 (2018) 1–18
- C. H. Chee, J. Jaafar, I. A. Aziz, M. H. Hasan, W. Yeoh, “Algorithms for frequent itemset mining: a literature review”, ArtifIntell Rev (2019) 52:2603–2621,
- Y. Wu, C. Zhu, Y. Li, L. Guo, X. Wu, “NetNCSP: Nonoverlapping closed sequential pattern mining”, Knowledge-Based Systems 196 (2020) 105812
- V. Dias, C. H. C. Teixeira, D. Guedes, W. Meira Jr., S. Parthasarathy, “Fractal: A General-Purpose Graph Patern Mining System”, SIGMOD ’19, June 30–July 5, 2019, Amsterdam, Netherlands, ACM
- W. Gan, J. C. W. Lin, P. F. Viger, H. C. Chao, P. S. Yu, “HUOPM: High Utility Occupancy Pattern Mining”, Journal of Latex Class Files, VOL. 14, NO. 8, AUGUST 2015
- W. Gan, J. C. W. Lin, P. F. Viger, H. C. Chao, P. S. Yu, “A Survey of Parallel Sequential Pattern Mining”, ACM Trans. Knowl. Discov. Data., Vol. 0, No. 1, Article 00. Publication date: August 2018.
- R. Bunker, K. Fujii, H. Hanada, I. Takeuchi, “Supervised sequential pattern mining of event sequences in sport to identify important patterns of play: An application to rugby union”, PLoS ONE 16(9): e0256329
- T. A. D. Lael, D. A. Pramudito, “Use of Data Mining for The Analysis of Consumer Purchase Patterns with The Fpgrowth Algorithm on Motor Spare Part Sales Transactions Data”, IAIC Transactions on Sustainable Digital Innovation (ITSDI), Vol. 4 No. 2 April 2023
- C. R. Wijesinghe, A. R. Weerasinghe, “Mining Frequent Patterns in Bioinformatics Workflows”, International Journal of Bioscience, Biochemistry and Bioinformatics, Volume 10, Number 4, October 2020
- Yang, W. Zhang, J. Wang, K. Yang, Y. Han, L. Zhang, “Review on the Application of Machine Learning Algorithms in the Sequence Data Mining of DNA”, Front. Bioeng. Biotechnol. 8:1032, 2020
- F. Wang, L. Liu, K. Li, N. Duić, M. S. khah, J. P. S. Catalão, “Impact Factors Analysis on the Probability Characterized Effects of Time of Use Demand Response Tariffs Using Association Rule Mining Method”, Energy Conversion and Management 197, 2019
- Y. Ali, A. Farooq, T. M. Alam, M. S. Farooq, M. J. Awan, T. I. Baig, “Detection of Schistosomiasis Factors Using Association Rule Mining”, IEEE access VOLUME 7, 2019
- M. H. Santoso, “Application of Association Rule Method Using Apriori Algorithm to Find Sales Patterns Case Study of IndomaretTanjungAnom”, brilliance research of artificial intelligence, Volume 1, Number 2, November 2021
- S. Das, A. Dutta, R. Avelar, K. Dixon, X. Sun, M. Jalayer, “Supervised association rules mining on pedestrian crashes in urban areas: identifying patterns for appropriate countermeasures”, International Journal of Urban Sciences, 2018