Enhancing Financial Fraud Detection using Temporal Patter Mining Technique

Ahmed Aziz[1,*], Sanjar Mirzaliev[1]

[1]Tashkent State University of Economics, Tashkent, Uzbekistan

Emails: a.mohamed@tsue.uz; sanjar2611@gmail.com

Abstract

Examining the temporal behavior of common patterns, obtaining appropriate clusters, and reducing the size of discovered patterns are three significant challenges in temporal data mining. Among the available methods, the constraint-based pattern mining approach has achieved remarkable progress in this domain. Apriori and Interleaved algorithms, which are both slow and outdated, are nonetheless used by present time-granularity pattern exploration approaches. To address these issues, we propose the Frequent Pattern Growth method with Special Constraints. The system incorporates a method for generating patterns on a regular basis. It mandates that transactional datasets adhere to complete and partial cyclic criteria. To locate all possible periodic patterns within the Spatio temporal database, we redefine the task as periodic pattern mining in this thesis. The proposed method makes use of a periodic pattern tree miner. To begin, the clustering method uses an innovative global pollination artificial fish swarm technique to create the most effective dense clusters.

Keywords: Fraud Detection; Financial Fraud Detection; Temporal Patter Mining Technique; Naïve Bayes; Radial Bias Neural Network

1. Introduction

Association Rule Mining (ARM) [1] is a data-mining topic that is currently generating a lot of controversy. Over the last two decades, storage devices have grown in capacity while shrinking in size and cost. Consequently, massive databases have replaced all other methods of data preservation and storage for customers ranging from sole proprietors to national governments. The goal of storing this information is to enable better decisions in the future via the discovery of connections and correlations among data points through further research. Discovering the links between different database elements is the current emphasis of data mining research. Consequently, finding frequent item sets is one of the most researched data mining fields. It is critical to consider methods to decrease candidate item sets and boost efficiency with frequent item sets while attempting to improve the data-mining algorithm. The generation of common item sets has been the subject of research in many different domains, including healthcare, banking, manufacturing, and telecommunications.

According to [2], the main motivation for researching Association Rule Mining comes from Market Basket Analysis. The purpose of this study is to identify any correlation between the purchases made by customers who make a single item buy at the two supermarkets. However, most applications rely on prior data analysis to predict future trends. A relatively recent field of research called temporal mining applies data mining techniques to temporal data in order to help with effective decision-making. Finding regular patterns is a critical step in mining data for linkages, correlations, and many other interesting interactions. Data mining processes such as data indexing, classification, and clustering get an advantage from it. Data mining research has mostly focused on frequent pattern mining due to its importance. The secret to accomplishing this is regularly mining item sets.

DOI: https://doi.org/10.54216/IJAACI.060206

Received: November 19, 2023 Revised: February 07, 2024 Accepted: July 09, 2024

1.1 Frequent Item set Mining

As a discipline of data mining, frequent item sets, mining has become more important and popular since it is fundamental to several crucial data mining tasks. A pattern is considered common when the number of occurrences of an item set in a dataset exceeds a support level that the user specifies. Consider the set me= {I1, I2, I3, Im} as an example of a collection of things. Take D, for example, as a database of transactions, where T is a subset of I since each transaction T adds a collection of objects. If the support-count of an item set I is more than or equal to the minimum support criterion, then it is deemed common according to [3]. Bread and butter, for example, would be categorized as a common item set in a dataset of purchase data due to the frequency with which they appear together.

An effective decision-making tool is the Apriori algorithm [4]. The Apriori method has a number of limitations, one of which being the time and energy needed to search the database and generate candidates. Because of this limitation, creating candidate Ck+1 takes more time and memory than usual. Moreover, when provided with inappropriate support levels and a large number of item sets [5] with extended durations, their performance suffers. Instead of checking if these candidates are frequent item sets after creating new candidates, this thesis proposes a better way for mining the association rules by making frequent k-item sets, which would lower the mining method's space and temporal complexity.

1.2 Temporal Data Mining

Time is an important concept in several subfields of database management systems. Rapid detection of viral infections in the bloodstream is a target of modern medical diagnostic tools. One thing that circuit-debugging algorithms consider is how long it takes for the charge on the capacitor to increase. Robot planning places a premium on job completion time. Applications that use data from the present, the past, and the future for purposes as if as prediction, planning, explanation, and learning are called temporal reasoning tasks. Systems for weather monitoring, demographic data and forest information all rely on these actions. According to [6], temporal data mining is the act of finding structures in temporal data by applying temporal patterns or models; it is a subset of knowledge discovery in temporal databases. The term "temporal data mining algorithm" refers to any method that can analyze time series data in order to find patterns or apply models to the data. The five main domains that temporal data mining comes under are pattern discovery, search and retrieval, classification, clustering, and prediction [7].

It is still up to the users to specify the minimum support count when mining databases using an Apriori-like algorithm. If the user is knowledgeable with the database, he may adjust the support range value according to his experience. On the other hand, inexperienced or uninformed users may find this task quite difficult. The current mining database does not allow users without database knowledge to modify the support range. Therefore, this research suggests and applies a User Preference Database method that keeps an eye on the database and helps new users when they need it most. The elimination of the need for transactional statistics is a key feature of this database.

2. Related Work

The process of clustering, an unsupervised learning technique widely used for many analytical tasks, is separating data into smaller, more manageable groups, or clusters [10]. Clustering methods may be broadly classified into two types: hierarchical and partitioned. By design, the hierarchical method allows items to belong to two clusters at once. Agglomerative and divisive hierarchical models exist. Agglomerative algorithms build individual clusters for each item before combining them into meaningful clusters. In contrast, after starting with all of the items, the division approach organizes the dataset into smaller groups.

One approach to collecting users' Spatio-temporal data and finding patterns in mobile phone communication was proposed in [11] as hierarchical agglomerative clustering (HAC). They discovered that subscribers often connect along a geographic proximity boundary by analyzing the paths that contact takes inside and between polygons. By seeing significant things and the connections between them, users may form social patterns that lead to meaningful interactions and the discovery of people with similar interests. Possible applications of these results include better distribution of information, better use of network resources, and city planning. A Spatio-temporal TOPOSCAN (ST-TOPOSCAN) method, which considers both the location and the time of vehicle trajectories, was proposed by [12] for evaluating the various user travel patterns and the underlying mechanisms for network assignment. The main objective of this study is to enhance the similarity within clusters using various criteria.

DOI: https://doi.org/10.54216/IJAACI.060206
Received: November 19, 2023 Revised: February 07, 2024 Accepted: July 09, 2024

The Spatio-temporal data mining problem was turned by [13] into an optimization problem. For typical Spatio-temporal data mining jobs, an optimization technique known as parameter-level stochastic gradient descent (plpSGD) was proposed. A major advantage of the plpSGD algorithm is that it improves speed while removing unnecessary processing. The method fixes the problem by updating just the parameters that have not achieved convergence at each iteration. Training the plpSGD model with the test data enhanced its ability to generalize.

A rule-mining framework that is associated with mobility from start to end may be used to infer the dynamics of urban mobility from the city's comprehensive traffic patterns [14]. It analyses the city's transportation demand, finds out how to predict transit requests and dynamic traffic based on mobility association principles, and examines improvements to the city's transportation network.

The Enhanced Krill Herd Algorithm was suggested by [15] for text document grouping. Partitioning the intricate clustering process into simpler ones allowed us to do things like use the particle swarm method for feature selection and the k-means approach for effectiveness assessment. We can see that the local search capacity has been improved from the experimental findings. Abualigah et al. created a hybrid clustering strategy by combining MMKHA with an upgraded krill herd algorithm. In this method, the parameter measurements used were F-measure, accuracy, entropy, precision, and ASDC. Research has shown that this technology outperforms other cutting-edge methods. Furthermore, this method failed to provide an effective and robust exploitation capacity. A novel feature selection method based on the particle swarm optimization algorithm was developed by [16] to improve document grouping. The accuracy and convergences of this approach were key determinants of performance. For the purpose of this evaluation, two case studies were utilized: the TRNSYS model and the energy plus model.

Several studies have examined how to effectively mine Spatio-temporal datasets for periodic sequences. A leading data mining approach is periodic sequence mining. In order to help with comprehension, the following section gives a brief synopsis of many significant study topics and pertinent literature. In order to make it easier to find different patterns from an initial Spatio-temporal GPS journey, [17] created a location matching pattern mining approach. However, this novel method uses a Hidden Markov Model to match sequences, which is how the trajectory patterns are matched. A case study centered on human paths was carried out using data derived from the spatial lives of the physical world. Furthermore, this method disregards the location matching mechanism's effectiveness and the mining of vast periodic patterns.

A permutation test based on the clustering approach was developed by [18] to find emerging patterns in Spatio-temporal datasets. We were able to find the most densely packed item and establish the cluster's importance using statistical data. In this experiment, a low variance test and a teat based on a rapid permutation test were both used. To begin, we used the local variance test to calculate the item densities. The many clusters were then generated using the cluster member permutation strategy, which exhibited a reduced homogeneity rate. The experiment proved that the novel method outperformed the competition. Use of a hierarchical approach allowed [19] to sidestep two significant restrictions: an unfamiliarity with hierarchical-based categorization and a lack of experience with trajectory sequencing.

Furthermore, a new clustering-based trajectory approach has been created that allows for the capture of Spatio-temporal data while also complying with several stringent constraints, such as time, direction, and speed. It was discovered that the novel hierarchical clustering algorithm for pattern mining was superior to a number of earlier approaches, including the Grid-based method and the kernel function method. When applied to a composite dataset, this approach was unable to define the multi-level period. In order to mine a huge number of new patterns on the internet all at once, [20] developed a dynamic frequent pattern tree.

A common dynamic tree for an unordered pattern and an ordered pattern were both explored further thereafter. One strong emergent pattern was produced from the combination of these approaches. This resulted in the development of a novel approach based on dynamic frequent pattern trees that achieves better performance than its forerunners.

3. Proposed Framework

This work makes use of the Reusable Time Ontology, a time ontology based on the idea of a timeline, which includes the Time Point and Time Interval classes as main components. Time units, which represent the degree of granularity, are another important notion. Further, down the hierarchy, there are two types of time intervals: Non-Convex Time Interval, which pertains to non-connected time intervals, and Convex Time Interval, which correlates to a connected interval on the timeline. The Regular Non-Convex Time Interval is a subclass of this type, with several convex time intervals representing frequent occurrences. Figure 1 is a simplified diagram showing this ontology.

DOI: https://doi.org/10.54216/IJAACI.060206
Received: November 19, 2023 Revised: February 07, 2024 Accepted: July 09, 2024

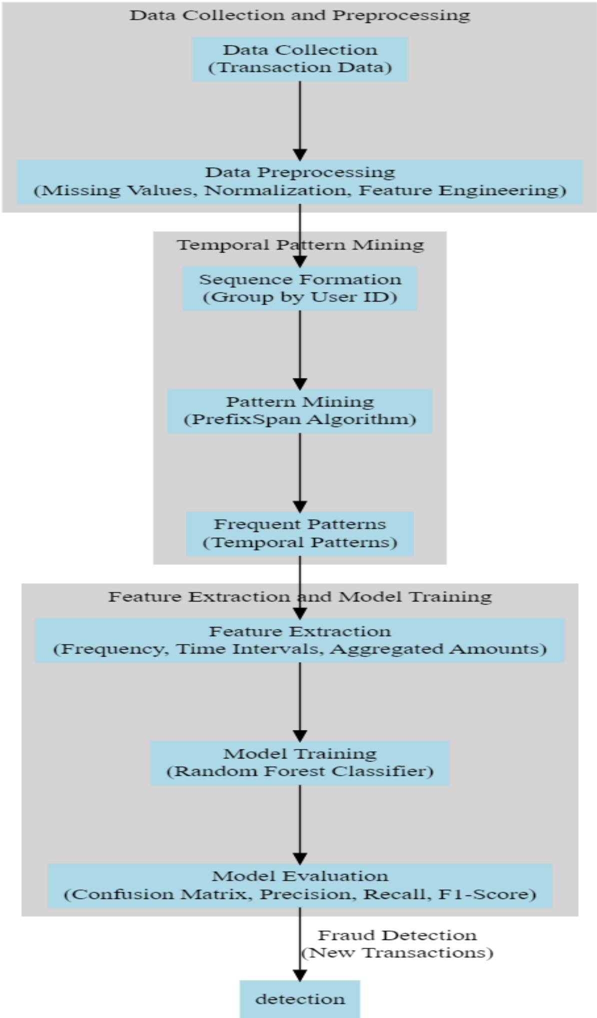

Figure 1. Diagram of time ontology based on reusability.

As we talk about time, there are a few fundamentals to keep in mind. The first is duration, the persistence of a trait across a certain time span. Second, order: at both intervals and time points, it shows whether one event occurs before or after another. The proximity or coincidence of two or more occurrences in time, regardless of their relative sequence, is called concurrency. However, there is also evolution, which explains how a pattern or event changes over time, and periodicity, which demonstrates how the same thing happens again.

3.1 System Architecture

Two issues arise from this: first, the methodologies do not adequately represent real-world circumstances, and second, time interval expressivity is absent. In addition, many approaches to more accurately portraying time were laid forth. For this purpose, the temporal ontology contributions of the D2PM and (TD) 2PaM algorithms are substantial.

- Identify sequences of transactions $T = \{t_1, t_2, \dots, t_n\}$ where each t_i has a timestamp and other attributes.
- Use algorithms like Prefix Span to mine frequent patterns $P = \{p_1, p_2, \dots, p_m\}$ from the transaction sequences.

For example, consider a simple sequence of transactions:

$$T = \{(t_1, A, 100, t1), (t_2, B, 200, t2), (t_3, A, 150, t3), (t_4, C, 300, t4)\} \quad (1)$$

A frequent pattern might be:

$$P = \{(A,*,t_i), (B,*,t_j)\} \quad (2)$$

DOI: https://doi.org/10.54216/IJAACI.060206
Received: November 19, 2023 Revised: February 07, 2024 Accepted: July 09, 2024

These methods make use of the interleaved algorithm in order to discover lifetime and periodic patterns. However, in terms of the amount of time required to build patterns, these techniques are ineffectual. These approaches make use of the Apriori method, which is a time-consuming method, in order to generate regular patterns.

3.2 User Interface

Because a user may only interact with the program via this user interface, the User Interface component of the proposed system is considered one of the most significant components in the application interface. In order to facilitate efficient communication with Decision Manager, this user interface delivers the necessary tools. Lastly, it is the responsibility of this user interface to collect the input from the user and then send it on to the Data Set, Rule Base, and Frequent Item set Generator components so that they can process the data.

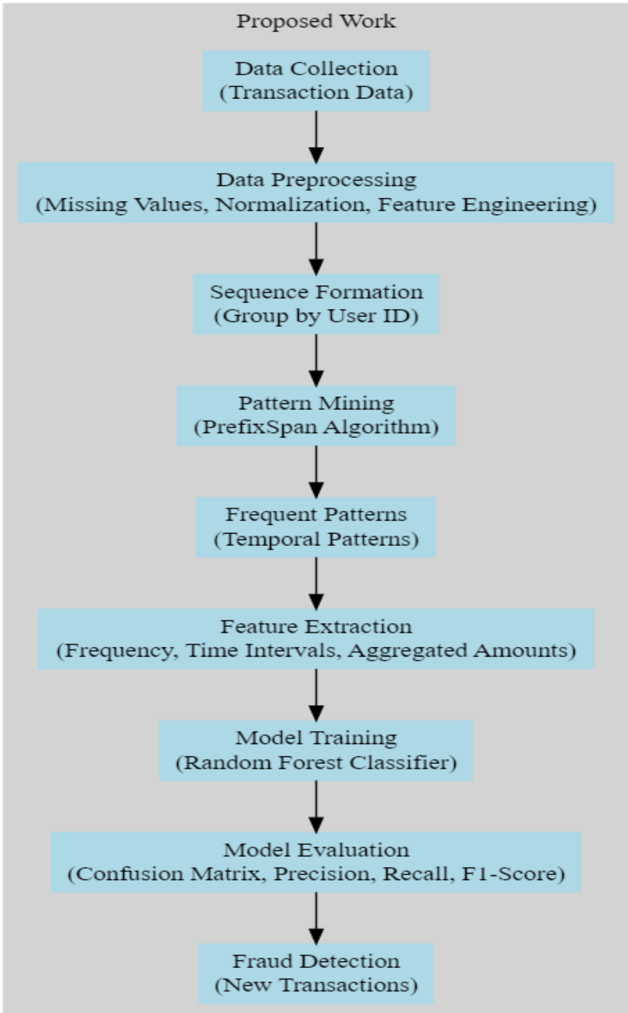

Figure 2. Flowchart of Proposed Work

3.3 Datasets

Within the scope of this study, the implementation of the Frequent Item sets generation has been carried out with the use of three distinct kinds of input data sets. To begin, the data set used by the University of California, Irvine (Frank et al. 2010 (Diabetic and ICU)) serves as the input data sets for evaluating the temporal frequent item set mining methods. Furthermore, the data set pertaining to travel and transactions is utilized as the input data set for confirming the development of the frequent item set via the utilization of the Bit Vector Mining methods. Finally yet importantly, the data from the mining logs are used as the input data sets in order to get the precise support count for the generation of frequent item sets.

DOI: https://doi.org/10.54216/IJAACI.060206
Received: November 19, 2023 Revised: February 07, 2024 Accepted: July 09, 2024

3.4 Rule Base

The rule base is capable of providing rule matching and rule firing techniques, so that accuracy level is increased.

- Extract features such as frequency of patterns, time intervals between patterns, and transaction amounts.
- For a pattern p occurring f_p times in a time window w :

$$\text{Frequency } = \frac{f_p}{w} \tag{3}$$

- Time interval between occurrences Δt :

$$\Delta t = t_{i+1} - t_i \tag{4}$$

- Aggregated transaction amounts for a pattern p :

$$\text{Amount}_p = \sum_{t \in p} \text{amount}(t) \tag{5}$$

The rules are stored and retrieved from the rule base, which consists of a set of IFS … THEN rules. This is an important subcomponent for the decision manager to make effective decisions.

3.5 Decision Manager

The person in charge of making decisions is really in charge of the system as a whole. A decision agent is utilized in this decision-making module. This decision agent is responsible for a variety of functions, one of which is the selection of an appropriate algorithm for the generation of frequent item sets from among the five proposed algorithms that are included in the Frequent Item set Generator. These algorithms are Temporal FP-Tree, Hashing Quadratic Probing, Space Preservation Mining; Enhanced Cluster based Bit Vector Association Rule, and Temporal Pattern Mining. Production rules are utilized in this decision-making module. For arriving at a conclusive choice, this module employs a collection of intelligent agents, namely the data set selection agent, the decision-making agent, and the action agent.

3.6 Frequent Item set Generator

It consists of two sub components namely Data Structure based and Bit Vector Mining based item set generator. The Data Structure component consists of two sub modules namely Temporal FP-Tree construction sub module and the Hashing Quadratic Probing (HQP) sub module. The Bit Vector Mining component consists of three sub modules namely Space Preservation Mining (SPM) sub module, Enhanced Cluster based Bit Vector Association Rule sub module and Temporal Pattern Mining sub module respectively. All these five sub modules are capable of generating frequent item sets from very large databases as well as from temporal databases.

3.6.1 Data Structure based Frequent Item set Generation

Frequent item sets are generated using the standard data structure component by two ways where, the first one uses the Temporal Frequent Pattern Tree algorithm and the second one uses hashing with Quadratic Probing for effective data mining. These mining methods are used to find 48 frequent item sets from temporal databases and conventional transaction databases respectively. Use machine-learning models like Logistic Regression, Random Forest, or Neural Networks.

- Define the feature vector X and target variable Y (fraud or non-fraud).
- Split the data into training and testing sets.
- Train the model using the training data:

$$\text{Model} : Y = f(X) \tag{6}$$

HBFI-QP does not use all the bins and hence, the phenomenon of primary clustering does not occur with quadratic probing. In this model, bit vectors are introduced to enhance the performance of rule mining. Based on bit vectors, three algorithms have been proposed in this work. Among them, the first method is the basic bit vector matrix representation; the second is the Enhanced Cluster based Bit Vector 49 Association Rule Mining and the third is the Bit Vector Temporal Pattern Mining algorithm.

3.7 User Profile Manager

If the minimum support value is very small then very large frequent items are discovered and hence, may find many irrelevant items. On the other hand, if the minimum support value is set with high values then the mining algorithm finds very few frequent item sets.

- Use the trained model to predict fraud on new transaction data.
- For a new transaction t_{new} with feature vector X_{new} :

DOI: https://doi.org/10.54216/IJAACI.060206
Received: November 19, 2023 Revised: February 07, 2024 Accepted: July 09, 2024

$$Y_{\text{pred}} = f(X_{\text{new}}) \quad (7)$$

This may lead to the omission of some important knowledge. Therefore, the setting of the minimum support is a difficult task in the mining world. Hence, this research work proposes, a user preference database that helps the user to fix the support value range.

4. Results and Discussion

We employed a systematic experimental setup to improve the detection of financial fraud with Temporal Pattern Mining (TPM). A synthetic dataset that simulated financial transactions was the primary data source. This dataset contained fields such as user ID, transaction amount, merchant, timestamp, and fraud labeled. To evaluate the model's efficacy, we conducted predictions on the test set and evaluated the results using a confusion matrix and classification metrics (precision, recall, F1-score):

Detailed breakdowns of true positives, true negatives, false positives, and false negatives were provided by the confusion matrix.

Classification Metrics: Precision, recall, and F1-score were calculated to assess the model's accuracy and potential to accurately identify fraudulent transactions. This experimental configuration guaranteed a comprehensive method for detecting financial deception through the application of temporal pattern mining and machine learning techniques.

Table 1: Toy dataset

1/1/2018	8 : 01	*x, y*
1/8/2018	18 : 30	*x, y, z*
1/14/2018	11 : 54	*x, y, z, u*
1/15/2018	16 : 34	*x, z*
1/21/2018	10 : 24	*y, u*
2/15//2018	17 : 33	*x, y, z*
2/13/2018	21 : 55	*z, u*
2/2/2018	18 : 27	*y, z, u*
2/27/2018	22 : 32	*y, z*
2/1/2018	13 : 34	*u, v*
3/15/2018	6 : 48	*x, y*
3/10/2018	11 : 23	*x, z*

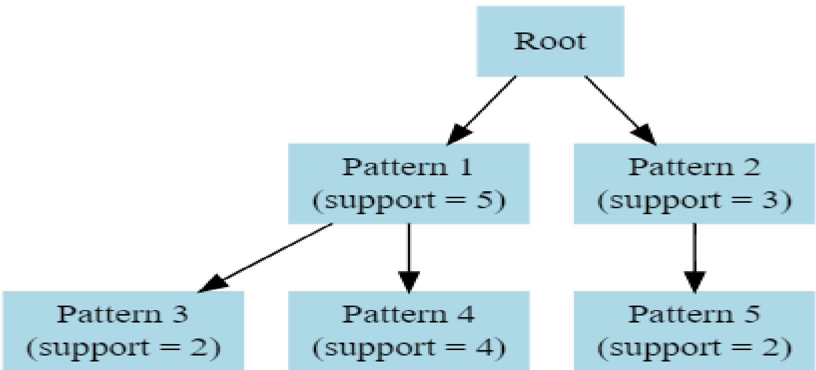

Figure 3. Frequent pattern tree for DB_0 (Jan-2018).

FP Growth processes to produce frequent item sets for each partition. Because the TD^2PM system employs the Apriori technique, the most time-consuming stage is generating frequent

DOI: https://doi.org/10.54216/IJAACI.060206
Received: November 19, 2023 Revised: February 07, 2024 Accepted: July 09, 2024

Table 2: Training and Testing items

Transaction No.	Transaction Date	Transaction Time	Items
1	1/1/2018	8 : 01	x, y
2	1/8/2018	18 : 30	x, y, z
3	1/14/2018	11 : 54	x, y, z, v
4	1/15/2018	16 : 34	x, z

Table 3: List of conditional patterns for partition DB_0 (Jan-2018)

z	$(x, y : 2)(x : 1)$	$(x : 3) \|z$
y	$(x : 3)$ empty	$(x : 3) \|y$
x	empty	empty

Conversely, the exploration space is diminished by the FP Growth approach. Consequently, we modify the TD2PM approach and replace the outdated process with the FP Growth one. The Inquire patterns approach then operates in two stages: first, it generates a comprehensive collection of candidate cycles, and then it eliminates extraneous cycles.

Table 4: collection of candidate cycles sample 1

u	$(z, y : 1)(z : 1)$	$(z : 2) \|u$
y	$(z : 3)$ empty	$(xz : 3) \|y$
z	empty	empty

Table 5: collection of candidate cycles sample 2

u	$(x, y : 1)(y, z : 1)$	$(y : 2) \|u$
z	$(y : 2)(x : 1)$	$(y : 2) \|z$
x	$(y : 2)$ empty	$(y : 2) \|x$

Table 6: collection of candidate cycles sample 3

v	$(x : 4)(x, u : 1)$	$(x : 5) \|u$
u	$(x : 4)(y : 1)$	Empty
y	$(x : 4)$ empty	$(x : 4) \|y$
x	empty	

Table 7: collection of candidate cycles sample 4

x	$[(p = 2, Y = 0)]$	Xz	$[(p = 3, Y = 0)]$
y	$[(p = 1, Y = 0), (p = 2, Y = 0), (p = 2, Y = 1)]$	Zu	$[(p = 2, Y = 0)]$
z	$[(p = 2, Y = 0), (p = 1, Y = 0), (p = 2, Y = 0)]$	Xv	$[(p = 4, Y = 0)]$
u	$[(p = 1, Y = 1)]$		
v	$[\text{Cycle}(p = 1, Y = 0)]$		

Table 8: collection of candidate cycles sample 5

x	$[(p = 1, Y = 0)]$	Zu	$[(p = 1, Y = 0)]$
y	$[(p = 2, Y = 0), (p = 1, Y = 0), (p = 2, Y = 0)]$	Xv	$[(p = 4, Y = 0)]$
u	$[(p = 1, Y 10), (p = 1, Y = 0), (p = 1, Y = 0)]$	Zu	$[(p = 1, Y = 0)]$
v	$[(p = 1, Y = 0)]$	Xu	$[(p = 4, Y = 0)]$

DOI: https://doi.org/10.54216/IJAACI.060206
Received: November 19, 2023 Revised: February 07, 2024 Accepted: July 09, 2024

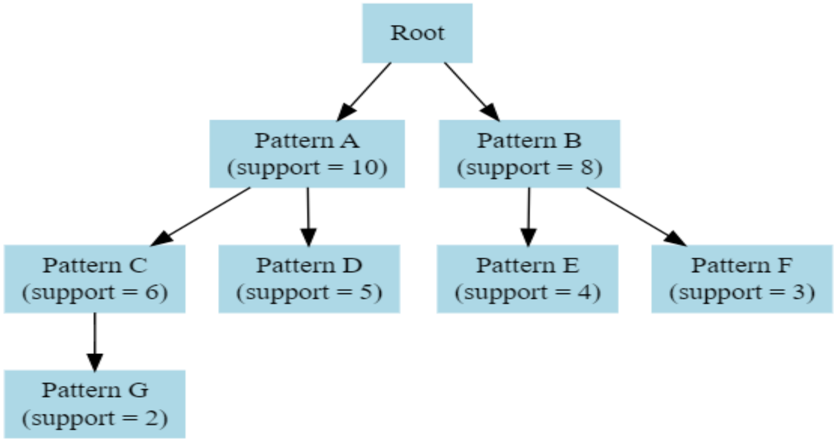

Figure 4. Frequent pattern tree for DB_1 (Feb-2018).

The stages for implementing the technique are as follows:

- The threshold for minimum support is set at two.
- The shortest and longest period length is set as 1 and 2 respectively.
- The sub-datasets and support are described in Table 2.

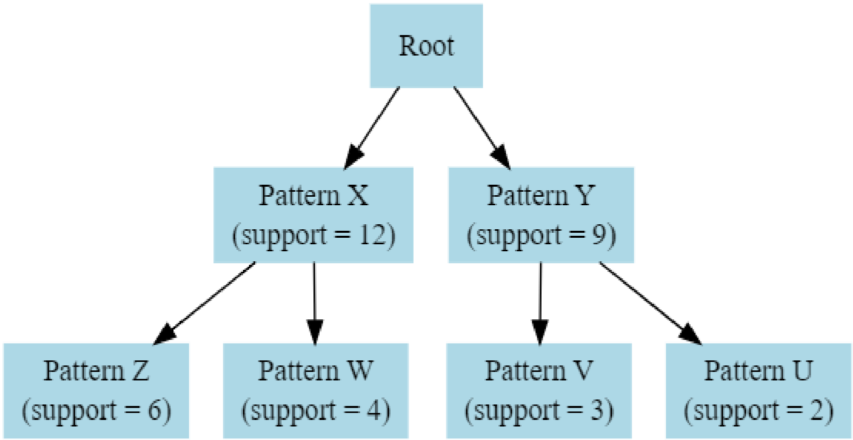

Figure 5. Frequent pattern tree for DB_2 (March-2018).

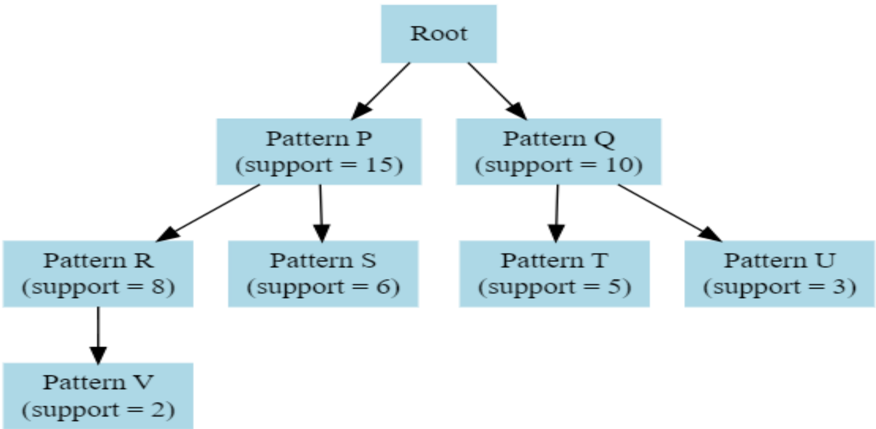

Fig. 6: Frequent pattern tree for DB_3 (April-2018).

DOI: https://doi.org/10.54216/IJAACI.060206
Received: November 19, 2023 Revised: February 07, 2024 Accepted: July 09, 2024

Approximately 2000 FM subscriber users have been included in this dataset over the past two years. Each row comprises the user's name, country name, date, composition, and vocalist. The algorithm's efficacy and performance, as well as the quantity of temporal patterns it generates, are subsequently assessed using 100,000 random recordings. The findings' efficacy is contrasted with that of the TD2PM framework.

The procedure is evaluated at two granularity levels: hour and month. The utmost duration period is parameterized as both a full and partial restriction. Figs. 6 and 7 illustrate the average time required to uncover each pattern, which is contingent upon the granularity (month and hour). The diagrams demonstrate that the time required establishing a pattern increases in tandem with the increase in the support threshold. The number of patterns decreases as the minimum support increases; second, the number of patterns increases by 1%; and third, the number of patterns substantially decreases as the minimum support increases by more than 10%.

The TD2PM algorithm is replaced with the FP Growth approach, which results in a decrease in the runtime time, as illustrated in Fig. 8. In comparison to the TD2PM framework, the total operating time was reduced by approximately 10% when FPGSC was employed. The FPGSC algorithm's operation time decreases as the minimal support threshold value increases, as it employs the FP Growth technique to identify frequent patterns.

After conducting a thorough analysis (views), the patterns that result can be observed in three different magnitudes. The first perspective of patterns is centered on the user, the second is centered on the vocalist, and the third is a combination of the two. Pattern P1 indicates that the individual in question listened to that particular performer on that website on a monthly basis. It indicates that the pattern is entirely cyclic. The period is one month, and the granularity level is one month. Conversely, Pattern 2 indicates that audiences prefer to listen to Gaga for a duration of two hours.

5. Conclusion

The primary objective is to optimize memory utilization by generating frequent item sets. It offers a capability to mitigate collisions during hashing, thereby enhancing the algorithm's time-based performance. This work proposes an enhanced Apriori algorithm that employs this hashing technique to facilitate the storage of obtained information. Additionally, the current algorithm minimizes the number of database scans and storage redundancy by pruning candidate item sets by infrequent item sets during the subset generation process. The provision of a facility for distributed data mining with intelligent agent-based communication is a potential solution for future work in this area. In addition, the association rules that are generated can be perpetually retained in a knowledge base and utilized to effectively perform the temporal reasoning task, which includes the analysis of the past, prediction of the future, planning, and learning activities. The primary focus of this work has been on the single-level association rule mining technique for temporal databases. Nevertheless, this research has the potential to be further developed to offer multi-level association rule mining techniques.

References

[1] Albashrawi, M. (2016). Detecting financial fraud using data mining techniques: A decade review from 2004 to 2015. *Journal of Data Science*, *14*(3), 553-569.

[2] Albashrawi, M. (2016). Detecting financial fraud using data mining techniques: A decade review from 2004 to 2015. *Journal of Data Science*, *14*(3), 553-569.

[3] Sharma, A., & Panigrahi, P. K. (2013). A review of financial accounting fraud detection based on data mining techniques. *arXiv preprint arXiv:1309.3944*.

[4] Phua, C., Lee, V., Smith, K., & Gayler, R. (2010). A comprehensive survey of data mining-based fraud detection research. *arXiv preprint arXiv:1009.6119*.

[5] Nami, S., & Shajari, M. (2018). Cost-sensitive payment card fraud detection based on dynamic random forest and k-nearest neighbors. *Expert Systems with Applications*, *110*, 381-392.

[6] Yang, W. S., & Hwang, S. Y. (2006). A process-mining framework for the detection of healthcare fraud and abuse. *Expert systems with Applications*, *31*(1), 56-68.

[7] Atluri, G., Karpatne, A., & Kumar, V. (2018). Spatio-temporal data mining: A survey of problems and methods. *ACM Computing Surveys (CSUR)*, *51*(4), 1-41.

DOI: https://doi.org/10.54216/IJAACI.060206

Received: November 19, 2023 Revised: February 07, 2024 Accepted: July 09, 2024

[8] Dal Pozzolo, A., Boracchi, G., Caelen, O., Alippi, C., & Bontempi, G. (2017). Credit card fraud detection: a realistic modeling and a novel learning strategy. *IEEE transactions on neural networks and learning systems*, *29*(8), 3784-3797.

[9] West, J., & Bhattacharya, M. (2016). Intelligent financial fraud detection: a comprehensive review. *Computers & security*, *57*, 47-66.

[10] Chang, R., Ghoniem, M., Kosara, R., Ribarsky, W., Yang, J., Suma, E., ... & Sudjianto, A. (2007, October). Wirevis: Visualization of categorical, time-varying data from financial transactions. In *2007 IEEE symposium on visual analytics science and technology* (pp. 155-162). IEEE.

[11] Hu, S., Zhang, X., Zhou, J., Ji, S., Yuan, J., Li, Z., ... & Fang, L. (2021, April). Turbo: Fraud detection in deposit-free leasing service via real-time behavior network mining. In *2021 IEEE 37th International Conference on Data Engineering (ICDE)* (pp. 2583-2594). IEEE.

[12] Zhang, X., Han, Y., Xu, W., & Wang, Q. (2021). HOBA: A novel feature engineering methodology for credit card fraud detection with a deep learning architecture. *Information Sciences*, *557*, 302-316.

[13] Hajek, P., & Henriques, R. (2017). Mining corporate annual reports for intelligent detection of financial statement fraud–A comparative study of machine learning methods. *Knowledge-Based Systems*, *128*, 139-152.

[14] Oluwafolake, A., & Solomon, O. A. (2017). A multi-algorithm data mining classification approach for bank fraudulent transactions. *African Journal of Mathematics and Computer Science Research*, *10*(1), 5-13.

[15] Vilalta, R., & Ma, S. (2002, December). Predicting rare events in temporal domains. In *2002 IEEE International Conference on Data Mining, 2002. Proceedings.* (pp. 474-481). IEEE.

[16] Wu, S. Y., & Chen, Y. L. (2007). Mining nonambiguous temporal patterns for interval-based events. *IEEE transactions on knowledge and data engineering*, *19*(6), 742-758.

[17] Vishal Dubey , Bhavya Takkar , Puneet Singh Lamba, Micro-Expression Recognition using 3D - CNN, Fusion: Practice and Applications, Vol. 1 , No. 1 , (2020) : 5-13 (Doi : https://doi.org/10.54216/FPA.010101)

[18] Surinder Kaur , Diksha Kumari , Vandana Kumari, Control of Enviornmental Parametrs in A Greenhouse, Fusion: Practice and Applications, Vol. 1 , No. 1 , (2020) : 14-21 (Doi : https://doi.org/10.54216/FPA.010102)

[19] Hardik Agarwal, Kanika Somani, Shivangi Sharma, Prerna Arora , Puneet Singh Lamba, Gopal Chaudhary*, Palmprint Recognition Using Fusion of Local Binary Pattern and Histogram of Oriented Gradients, Fusion: Practice and Applications, Vol. 1 , No. 1 , (2020) : 22-31 (Doi : https://doi.org/10.54216/FPA.010103)

[20] Aditya Sharma , Aditya Vats , Shiv Shankar Dash , Surinder Kaur, Artificial Intelligence enabled virtual sixth sense application for the disabled, Fusion: Practice and Applications, Vol. 1 , No. 1 , (2020) : 32-39 (Doi : https://doi.org/10.54216/FPA.010104)

DOI: https://doi.org/10.54216/IJAACI.060206

Received: November 19, 2023 Revised: February 07, 2024 Accepted: July 09, 2024