International Journal of Science and Research Archive
eISSN: 2582-8185
Cross Ref DOI: 10.30574/ijsra
Journal homepage: https://ijsra.net/

(Review Article)

Discovering temporal frequent patterns: An optimized approach

Sheel Shalini * and Sweta Kumari

Department of Computer Science and Engineering, Birla Institute of Technology Mesra, Patna, India.

International Journal of Science and Research Archive, 2024, 12(02), 1273–1278

Publication history: Received on 30 March 2024; revised on 23 July 2024; accepted on 26 July 2024

Article DOI: https://doi.org/10.30574/ijsra.2024.12.2.1366

Abstract

Temporal frequent pattern mining is a critical area of data mining that focuses on identifying recurring set of events over time. Traditional methods often face challenges related to scalability, data complexity, and noise, necessitating the development of optimized approaches. This survey reviews the existing literature on temporal frequent pattern mining, highlights the limitations of traditional techniques, and presents recent advancements in optimized methods. The paper also discusses future directions and applications across various domains.

Keywords: Temporal Frequent pattern Mining; Optimized Approaches; Parallel and Distributed Computing; Dynamic Time Window; Pattern Pruning Technique; Recent Advancements

1. Introduction

Temporal association rule mining is a data mining technique [1] that discovers association rules within a specified time period. Temporal association rules are of the form: $(X \rightarrow Y)^{TP}$ where, X and Y are items and TP is specified time period. The rules generated by mining a temporal database is valuable for identifying time-dependent associations between items [2] and for making decisions. The abundance of temporal data has spurred researchers to uncover various types of time-variant patterns and regularities hidden within these databases.

Frequent pattern mining plays a vital role in temporal association rule mining as frequent patterns identified within temporal databases serve as the basis for generating temporal association rules. Temporal frequent pattern mining is first step in temporal association rule mining[3].

Temporal frequent patterns involve sequences of events that occur frequently within a given time frame. These patterns can help predict future events, detect anomalies, and provide a deeper understanding of temporal relationships in data. For instance, in retail, discovering that certain products are frequently bought together during specific seasons can inform inventory and marketing strategies.

Temporal frequent pattern mining involves extracting patterns from time-ordered data, which can reveal valuable insights for decision-making in areas such as finance[4], healthcare[5], retail[6], and more. Traditional techniques, such as Apriori [7] and FP-Growth [8] based algorithms for temporal frequent pattern mining often struggle with the volume and complexity of temporal data. This survey aims to provide an overview of optimized approaches that address these challenges, enhancing the efficiency and scalability of pattern discovery.

* Corresponding author: Sheel Shalini

1.1. Challenges in Temporal Pattern Mining

Effective temporal pattern mining must address these challenges by leveraging advanced techniques such as scalable data structures, parallel processing, and robust statistical methods to manage data volume and unravel the intricate patterns hidden within complex temporal datasets.

- **Data Volume and Complexity:** Temporal data can be vast and intricate, requiring algorithms to handle large-scale processing efficiently. Temporal pattern mining involves analyzing time-ordered data to discover patterns that describe the temporal relationships between events. This field faces significant challenges related to the volume often contain millions or billions of time-stamped events especially in fields like finance, healthcare, and telecommunications. and complexity of the data. Handling, storing, and processing such large volumes of data require substantial computational resources and efficient algorithms to ensure timely analysis.
- **Time Constraints:** Temporal patterns are sensitive to the time window considered. Patterns may vary significantly based on the chosen time window necessitating methods that can adapt to varying time frames.
- **Scalability:** Traditional algorithms may not scale well with increasing data size and complexity. Algorithms must scale to accommodate increasing data sizes without significant performance degradation.
- **Noise and Incomplete Data:** Large temporal data often contains noise and missing values, complicating pattern discovery.

2. Optimized Approaches

Optimized approaches are used in various fields and applications to achieve better performance, efficiency, and effectiveness. Optimization helps in utilizing resources (time, memory, computational power, etc.) more efficiently. This is crucial in environments with limited resources, such as embedded systems or mobile devices. By optimizing processes or systems, costs associated with energy consumption, materials, and labor can be reduced. Optimized approaches often lead to faster and more responsive systems.

2.1. Parallel and Distributed Computing

Parallel and distributed computing techniques are helpful for addressing the challenges in temporal pattern mining, where the goal is to discover meaningful patterns from time-stamped data. Employing techniques such as parallel computing is essential to effectively manage the heightened time and space complexities by distributing the workload across multiple processors [9]. These computing paradigms are specifically utilized in:

- **Handling Large-scale Data**: Temporal pattern mining often involves large volumes of time-series data, which can be distributed across multiple machines or processed concurrently on multicore processors. Parallel computing allows for efficient handling of these large datasets by dividing the workload among multiple computing units.
- **Parallel Algorithms**: Algorithms for temporal pattern mining, such as sequence mining or frequent pattern mining, can be parallelized to improve efficiency[10]. For example, algorithms like PTP utilizes multithreading on a frequent temporal pattern tree where each branch is processed in parallel. It can be adapted to run concurrently on subsets of data or time intervals, speeding up the discovery of temporal patterns.
- **Distributed Frameworks**: Distributed computing frameworks like Apache Spark or Hadoop provide tools and libraries that facilitate the implementation of temporal pattern mining algorithms across clusters of machines. These frameworks enable data partitioning, parallel execution of tasks, and fault tolerance, which are critical for handling large-scale temporal data effectively. Distributed computing frameworks offer built-in fault tolerance mechanisms, which ensure that temporal pattern mining tasks can continue execution even if some nodes in the cluster fail. Distributed Hierarchical Pattern Graph TPM (DHPG-TPM) that supports large-scale TPM using the leading distributed platform Apache Spark [11].

By distributing the workload across multiple nodes or processors, parallel and distributed computing architectures can achieve significant speedups compared to single-machine processing. This is particularly beneficial for iterative algorithms used in temporal pattern mining.

2.2. Advanced Data Structures

In temporal frequent pattern mining, where the goal is to discover patterns that occur frequently over time, advanced data structures play a crucial role in efficiently storing and accessing temporal data. There are various advanced data structures for temporal pattern mining. Each of these advanced data structures offers unique advantages for managing and analyzing temporal data in the context of frequent pattern mining. Their selection depends on specific requirements such as the type of temporal data, the frequency of patterns being mined, and the efficiency needed for querying and updating operations. Integrating these structures with appropriate algorithms [12] for pattern mining can significantly enhance the effectiveness and performance of temporal frequent pattern mining systems.

Here are some advanced data structures commonly used for optimization:

- **Interval Trees**: Interval trees are useful for storing and querying intervals, which are prevalent in temporal data where events or patterns occur over time intervals [13]. They facilitate efficient retrieval of intervals that overlap with a given query interval, which is essential for tasks like finding frequent patterns that occur within specific time spans.
- **Segment Trees**: Segment trees are versatile for querying and updating intervals or segments of data. They can be adapted to handle temporal data by storing intervals or timestamps, supporting operations like range queries (e.g., finding frequent patterns within a range of timestamps) and updates (e.g., adding new temporal events) [14][15].
- **Suffix Trees** and **Suffix Arrays:** They facilitate efficient pattern matching over time sequences, helping to identify recurring patterns or sequences of events that occur frequently within specified time intervals [16].
- **Trie Structures**: Tries (prefix trees) are useful for storing and retrieving sequences of elements efficiently [17]. In temporal frequent pattern mining, tries can be employed to store temporal sequences (e.g., sequences of events over time) and facilitate operations such as identifying frequent subsequences or patterns.
- **Compressed Data Structures**: These structures aim to reduce memory usage and improve query performance by compressing data while maintaining efficient access [18]. In temporal frequent pattern mining, compressed data structures can be applied to store and query large volumes of temporal data, supporting tasks such as identifying frequent patterns or trends over time efficiently [19].

2.3. Dynamic Time Windows

Dynamic Time Windowing is a crucial concept in temporal data mining, especially for tasks like frequent pattern mining where patterns must be discovered within specific time intervals. This approach addresses the challenge of analyzing time-varying data streams or sequences where patterns may evolve over time. In temporal frequent pattern mining, the concept of a dynamic time window refers to a sliding or variable-sized window that moves over the timeline of events or transactions. This dynamic window allows the mining algorithm to focus on patterns that occur within a specific time frame or interval, adjusting as new data arrives or as patterns change over time [20].

Several techniques have been developed to implement Dynamic Time Windowing effectively:

- **Sliding Windows:** A fixed-size window slides over the data stream or sequence at regular intervals. This approach is straightforward but may not adapt well to varying data densities or event frequencies [21].
- **Variable-sized Windows:** The size of the window adjusts dynamically based on criteria such as event frequency, data density, or pattern characteristics. This flexibility allows for more precise pattern detection [22].
- **Incremental Updating:** As new data arrives, algorithms update the window position and size incrementally, ensuring continuous pattern discovery without reprocessing entire data sets [23].

2.4. Pattern Pruning Techniques

Pattern pruning techniques in temporal frequent pattern mining are essential for improving the efficiency and effectiveness of pattern discovery algorithms. These techniques focus on reducing the search space by eliminating irrelevant or redundant patterns, thereby enhancing the scalability and interpretability of mining results. In temporal data mining, especially in scenarios involving sequences of events with timestamps, the volume of potential patterns can be immense. Pattern pruning techniques aim to mitigate this issue by selectively removing patterns that do not meet certain criteria or are less relevant to the analysis task [20]. This process is crucial for reducing computational complexity, improving interpretability, enhancing pattern quality, etc.

Several techniques are used to prune patterns in temporal frequent pattern mining:

- **Minimum Support Thresholding:** This is a fundamental pruning technique where patterns that do not meet a minimum support threshold (i.e., frequency of occurrence) are discarded. For temporal patterns, this threshold can be adjusted dynamically based on time intervals or event frequencies [20].
- **Pattern Growth Pruning:** In algorithms like FP-Growth for temporal data, patterns are grown incrementally [24]. Pruning can occur during the growth process by discarding candidate patterns early if they do not contribute significantly to frequent pattern discovery.
- **Temporal Constraints:** Pruning based on temporal constraints such as maximum gap between events or minimum duration of patterns helps refine the discovered patterns to adhere more closely to temporal characteristics of interest[25].
- **Interestingness Measures:** Applying measures such as lift, confidence, or novelty to assess the significance of patterns and pruning those that do not meet certain interestingness criteria [26].

3. Recent Advancements in Temporal Frequent Pattern Mining

Recent advancements in temporal pattern mining have introduced various innovative techniques and frameworks and significantly improved the ability to analyze and extract meaningful patterns from time-dependent data. Key developments include:

- **Temporal Network Analysis**: Advances in mining temporal networks have enabled better understanding of dynamic systems, such as social interactions and communication patterns, by identifying temporal motifs and centrality measures [27].
- **Extended Lists Algorithm**: A faster algorithm for mining frequent temporal patterns has been developed, significantly outperforming existing methods. This approach uses extended lists to enhance efficiency, making it particularly effective on both real-life and random data [28].
- **Temporal Inter-object Pattern Mining**: This method focuses on identifying patterns between different objects over time in multivariate time series. It provides more informative results by incorporating explicit temporal information, which is crucial for applications like stock market analysis and healthcare monitoring[29].
- **Shapelet-based Approaches**: Shapelet-based methods identify characteristic subsequences (shapelets) in time series data that are critical for understanding temporal patterns. These approaches enhance the interpretability and accuracy of the mined patterns, which is beneficial for domains like medical diagnosis and climate science [29].
- **Generative Adversarial Networks (GANs)**: The application of GANs in temporal pattern mining has shown promise, particularly in handling multivariate time series. GANs help address the divergence of subsequences from the original data, thereby improving the quality of the discovered patterns [28].
- **IoT and Sensor Data**: The proliferation of IoT devices has led to the development of efficient and distributed temporal pattern mining algorithms to handle the massive influx of time-series data from sensors [30].
- **Deep Learning Techniques**: The incorporation of deep learning has improved the extraction of complex temporal patterns, allowing for more accurate predictions and anomaly detection in various domains [31].

4. Future Research in Temporal Frequent Pattern Mining

Future research in temporal frequent pattern mining is poised to explore several promising directions, driven by the increasing complexity of time-series data and the growing demand for real-time, interpretable, and actionable insights. Some future direction for temporal frequent pattern mining are:

- A hybrid model Combining traditional pattern mining techniques with deep learning models, such as recurrent neural networks (RNNs) and long short-term memory (LSTM) networks, to capture complex temporal dependencies.
- Creating algorithms can adapt to changes in the data distribution over time, ensuring that the mined patterns remain relevant and accurate.
- Developing algorithms for real-time analysis of streaming data for enabling immediate insights and decision-making in dynamic environments.
- Combining multiple optimized techniques, such as integrating machine learning with advanced data structures, for enhanced performance.

5. Conclusion

Temporal frequent pattern mining is a vital technique for uncovering valuable insights from time-ordered data. While traditional methods have laid the groundwork, optimized approaches are essential for addressing the challenges posed by large-scale, complex, and noisy datasets. By leveraging parallel computing, advanced data structures, dynamic time windows, and integrating machine learning, these optimized methods significantly enhance the efficiency and accuracy of temporal pattern discovery. Continued research and development in this area will unlock new opportunities and applications across various domains.

Compliance with ethical standards

Disclosure of conflict of interest

No conflict of interest to be disclosed.

References

[1] J. M. Ale and G. H. Rossi, “An approach to discovering temporal association rules,” in Proc. SAC, Como, Italy, 2000, pp. 294–300, DOI: 10.1145/335603.335770.

[2] Segura-Delgado, A., Gacto, M. J., Alcalá, R., & Alcalá-Fdez, J. (2020). Temporal association rule mining: An overview considering the time variable as an integral or implied component. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 10(4), e1367.

[3] Tan T. F, Wang Q, G, Phang T. H, Li X, Huang J, Zhang D. Temporal Association Rule Mining. In: He X. et al. (eds) Intell. Sci. Big Data Eng., Springer, Cham, 2015, pp. 247-257.

[4] Nair, B. B., Mohandas, V. P., Nayanar, N., Teja, E. S. R., Vigneshwari, S., & Teja, K. V. N. S. (2015). A stock trading recommender system based on temporal association rule mining. SAGE Open, 5(2), 2158244015579941.

[5] Concaro, S., Sacchi, L., Cerra, C., Fratino, P., & Bellazzi, R. (2009). Mining healthcare data with temporal association rules: Improvements and assessment for a practical use. In Artificial Intelligence in Medicine: 12th Conference on Artificial Intelligence in Medicine, AIME 2009, Verona, Italy, July 18-22, 2009. Proceedings 12 (pp. 16-25). Springer Berlin Heidelberg.

[6] Guidotti R, Gabrielli L, Monreale A, Pedreschi D, Giannotti F. Discovering temporal regularities in retail customers' shopping behavior. EPJ Data Sci., 2018, 7, 6.

[7] Agrawal, R., & Srikant, R. (1994). Fast algorithms for mining association rules in large databases. Proceedings of the 20th International Conference on Very Large Data Bases (VLDB).

[8] Han, J., Pei, J., & Yin, Y. (2000). Mining frequent patterns without candidate generation. Proceedings of the 2000 ACM SIGMOD International Conference on Management of Data.

[9] Ho, V. L., Ho, N., Pedersen, T. B., & Papapetrou, P. (2023). Efficient Generalized Temporal Pattern Mining in Big Time Series Using Mutual Information. arXiv preprint arXiv:2306.10994.

[10] Vu, N.T., Vo, C. (2021). A Parallelized Frequent Temporal Pattern Mining Algorithm on a Time Series Database. In: Nguyen, N.T., Chittayasothorn, S., Niyato, D., Trawiński, B. (eds) Intelligent Information and Database Systems. ACIIDS 2021. Lecture Notes in Computer Science(), vol 12672. Springer, Cham. https://doi.org/10.1007/978-3-030-73280-6_7.

[11] N. Ho, V. L. Ho, T. Bach Pedersen and M. Vu, "Efficient and Distributed Temporal Pattern Mining," 2021 IEEE International Conference on Big Data (Big Data), Orlando, FL, USA, 2021, pp. 335-343, doi: 10.1109/BigData52589.2021.9671753.

[12] Dhakshayani, J., Sivasathya, S., Sharmiladevi, S., Sophie, S.L. (2023). Mining Frequent Patterns from Temporal Dataset Using Backtracking Search Tree of GenMax Algorithm. In: Bhattacharyya, S., Banerjee, J.S., Köppen, M. (eds) Human-Centric Smart Computing. Smart Innovation, Systems and Technologies, vol 316. Springer, Singapore. https://doi.org/10.1007/978-981-19-5403-0_9

[13] Mordvanyuk, N., Lopez, B., & Bifet, A. (2021). vertTIRP: Robust and efficient vertical frequent time interval-related pattern mining. Expert Systems with Applications, 168, 114276.

[14] Jin, L., Lee, Y., Seo, S., Ryu, K.H. (2006). Discovery of Temporal Frequent Patterns Using TFP-Tree. In: Yu, J.X., Kitsuregawa, M., Leong, H.V. (eds) Advances in Web-Age Information Management. WAIM 2006. Lecture Notes in Computer Science, vol 4016. Springer, Berlin, Heidelberg. https://doi.org/10.1007/11775300_30

[15] Jazayeri, A., & Yang, C. C. (2023). Frequent pattern mining in continuous-time temporal networks. IEEE transactions on pattern analysis and machine intelligence.

[16] Jia, L., Zhou, C., Wang, Z., & Xu, X. (2005, August). SuffixMiner: Efficiently mining frequent itemsets in data streams by suffix-forest. In International Conference on Fuzzy Systems and Knowledge Discovery (pp. 592-595). Berlin, Heidelberg: Springer Berlin Heidelberg

[17] Hosseininasab, A., van Hoeve, W. J., & Cire, A. A. (2022). Memory-Efficient Sequential Pattern Mining with Hybrid Tries. arXiv preprint arXiv:2202.06834.

[18] Jayasankar, U., Thirumal, V., & Ponnurangam, D. (2021). A survey on data compression techniques: From the perspective of data quality, coding schemes, data type and applications. Journal of King Saud University-Computer and Information Sciences, 33(2), 119-140.

[19] Liu, J., Ye, Z., Yang, X., Wang, X., Shen, L., & Jiang, X. (2022). Efficient strategies for incremental mining of frequent closed itemsets over data streams. Expert Systems with Applications, 191, 116220.

[20] Shalini, S., & Lal, K. (2019). Mining Changes in Temporal Patterns in Latest Time Window for Knowledge Discovery. Journal of Information & Knowledge Management, 18(03), 1950028.

[21] Tanbeer, S. K., Ahmed, C. F., Jeong, B. S., & Lee, Y. K. (2009). Sliding window-based frequent pattern mining over data streams. Information sciences, 179(22), 3843-3865.

[22] Li, H., & Wang, L. (2017, May). A variable size sliding window based frequent itemsets mining algorithm in data stream. In AIP Conference Proceedings (Vol. 1839, No. 1). AIP Publishing.

[23] Lee, C. H., Lin, C. R., & Chen, M. S. (2001, October). Sliding-window filtering: an efficient algorithm for incremental mining. In Proceedings of the tenth international conference on Information and knowledge management (pp. 263-270).

[24] Cheng, H., Han, J. (2016). Pattern-Growth Methods. In: Liu, L., Özsu, M. (eds) Encyclopedia of Database Systems. Springer, New York, NY. https://doi.org/10.1007/978-1-4899-7993-3_263-2

[25] A. Omari, "A new temporal measure for interesting frequent itemset mining," 2010 2nd IEEE International Conference on Information Management and Engineering, Chengdu, China, 2010, pp. 425-429, doi: 10.1109/ICIME.2010.5477848.

[26] Zhang, Y., & Paquette, L. (2020). An effect-size-based temporal interestingness metric for sequential pattern mining. learning, 4, 16.

[27] Jazayeri, A., & Yang, C. C. (2023). Frequent pattern mining in continuous-time temporal networks. IEEE transactions on pattern analysis and machine intelligence.

[28] Kocheturov, A., Pardalos, P.M. (2018). Frequent Temporal Pattern Mining with Extended Lists. In: Mondaini, R. (eds) Trends in Biomathematics: Modeling, Optimization and Computational Problems. Springer, Cham. https://doi.org/10.1007/978-3-319-91092-5_16

[29] Vu, N.T., Chau, V.T.N. (2014). Frequent Temporal Inter-object Pattern Mining in Time Series. In: Huynh, V., Denoeux, T., Tran, D., Le, A., Pham, S. (eds) Knowledge and Systems Engineering. Advances in Intelligent Systems and Computing, vol 244. Springer, Cham. https://doi.org/10.1007/978-3-319-02741-8_15

[30] Braun, P., Cuzzocrea, A., Leung, C. K., Pazdor, A. G., Tanbeer, S. K., & Grasso, G. M. (2018). An innovative framework for supporting frequent pattern mining problems in IoT environments. In Computational Science and Its Applications–ICCSA 2018: 18th International Conference, Melbourne, VIC, Australia, July 2-5, 2018, Proceedings, Part V 18 (pp. 642-657). Springer International Publishing.

[31] Jamshed, A., Mallick, B. & Kumar, P. Deep learning-based sequential pattern mining for progressive database. Soft Computing 24, 17233–17246 (2020). https://doi.org/10.1007/s00500-020-05015-2