Review Article

ACCENTS Transactions on Information Security, Vol 8(29)
ISSN (Online): 2455-7196
http://dx.doi.org/10.19101/TIS.2023.829001

Innovations in frequent itemset mining: challenges and opportunities

Abhilash Behera* **and Md Zuber**
School of Computer Science and Engineering, Madhyanchal Professional University, Bhopal, India

Received: 20-December-2022; Revised: 15-January-2023; Accepted: 20-January-2023

Abstract
The era of big data, characterized by vast and complex datasets, has prompted the need for advanced data mining techniques. Frequent itemset mining, a fundamental method in data mining, plays a pivotal role in uncovering hidden knowledge and patterns. However, it faces challenges in scalability, adaptability to uncertainty, and the need to consider rare and closed itemsets. This paper reviews recent advancements in frequent itemset mining, focusing on innovative approaches introduced in 2022 and 2023. These advances address dynamic database updates, efficient fault prediction, scalability issues, quantitative pattern mining, utility-based approaches, mining rare itemsets, real-time decision-making, and uncertain frequent itemset mining. While these studies offer valuable solutions, they also present challenges related to scalability, adaptability, and performance. Future research should refine these methods to meet evolving data mining demands.

Keywords
Frequent itemset mining, Data Mining, Decision making, Scalability.

1.Introduction
In today's digital age, the ubiquity of advanced computer technology and the pervasive use of Internet products have propelled us into the era of big data. The consequences of this transition are monumental, as we now find ourselves inundated with vast and complex datasets that contain an abundance of valuable information. Consequently, researchers and data scientists are increasingly driven to develop advanced data mining techniques to extract meaningful insights from this ocean of data.

One of the pivotal techniques that has played a crucial role in uncovering hidden knowledge and patterns within these expansive datasets is frequent itemset mining [1-3]. This method, also known as frequent pattern mining, is a foundational concept in the realm of data mining [4, 5]. It has found widespread applications in various domains, including Internet applications, financial analysis, healthcare, and many more [6-8]. At its core, frequent itemset mining aims to identify sets of items that regularly co-occur in each dataset [9, 10].

The significance of frequent itemset mining lies in its capacity to unearth associations and patterns within the data [11, 12]. These patterns are essential for making data-driven decisions, identifying trends, and gaining deeper insights into complex datasets [13, 14]. Consequently, researchers and practitioners have harnessed this technique to extract valuable information and drive data-driven processes in numerous industries and applications [15, 16].

While frequent itemset mining has proven to be a reliable and invaluable tool, it is not without its challenges and limitations. Over the years, researchers have recognized the need to address these issues to enhance the efficiency and applicability of this technique in the context of big data and modern data analytics [17, 18].

Considering the ever-expanding volume and complexity of data, the scalability of frequent itemset mining has emerged as a critical challenge [19, 20]. Traditional methods may struggle to handle the sheer size of datasets and the intricacies of contemporary data sources. This necessitates the development of more scalable algorithms and strategies that can efficiently process and analyze data on a grand scale [21, 22].

*Author for correspondence

Furthermore, in today's data landscape, where uncertainty is prevalent, the conventional techniques for frequent itemset mining may fall short. Modern data sources often contain probabilistic elements, making it imperative to adapt existing mining methods to account for this uncertainty. As a result, researchers have been exploring innovative approaches that incorporate probabilistic thresholds and more comprehensive support measures.

Moreover, as data mining evolves, there is a growing recognition of the importance of mining rare itemsets and closed itemsets in addition to frequent ones [23-25]. These less common but often valuable patterns have garnered attention, as they can provide unique and critical insights into the data [23-25].

Frequent itemset mining remains a vital and evolving field in the realm of data mining. The era of big data has ushered in both unprecedented opportunities and formidable challenges, making it imperative for researchers to continually adapt and innovate in their quest to extract valuable insights from the vast and intricate datasets at our disposal. This paper aims to review and consolidate recent advancements in frequent itemset mining, shedding light on how innovative approaches are shaping the field and providing valuable guidance for future research and applications in this exciting and dynamic domain. Figure 1 illustrates the application of frequent itemset mining, highlighting several major areas where this technique can be effectively utilized.

This paper is structured as follows: Section 2 presents the literature review. Section 3 explores the discussion and analysis. Finally, Section 4 provides the concluding remarks.

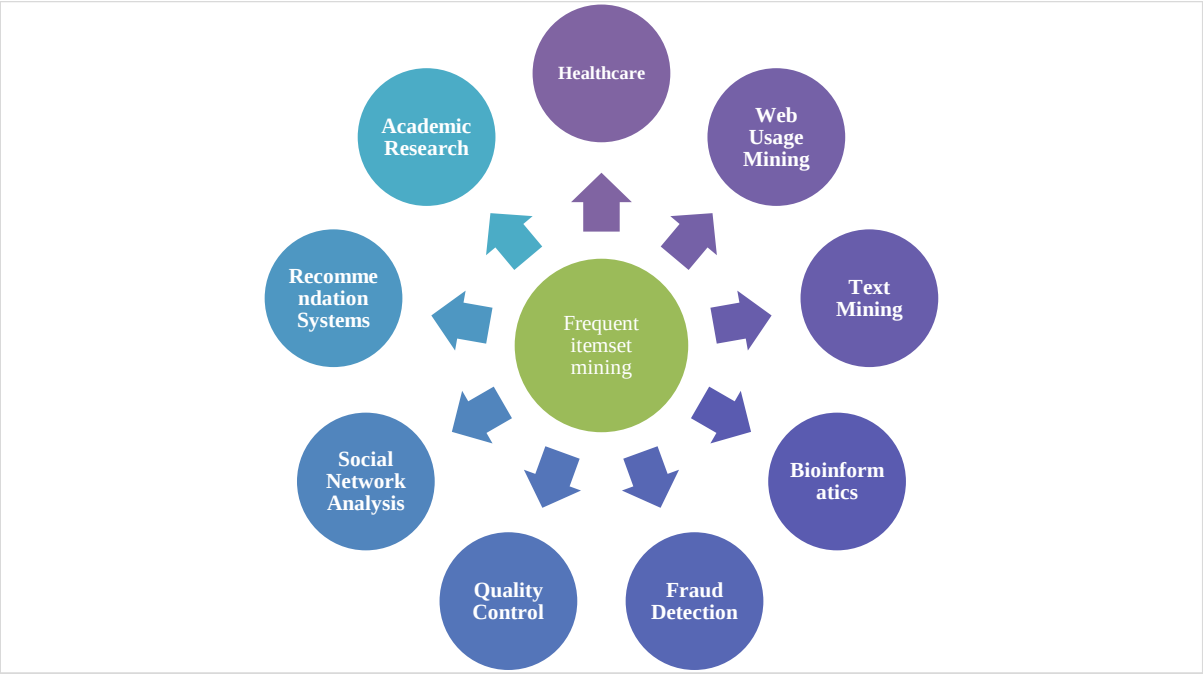

Figure 1 Applications of frequent itemset mining

2.Literature review

In 2022, an incremental method was proposed by Hong et al [26]. It extended fuzzy utility mining, using a fast-update strategy, to adapt to dynamic databases. Experimental results demonstrated its superiority over the batch-based approach, ensuring correct knowledge updates.

In 2022, He and Lu [27] employed frequent itemset mining to analyze equipment fault associations in complex power grid structures. Correlated fault features were selected using mutual information, and a model assessed potential relationships. Their method outperformed other algorithms in training time and accuracy, aiding power grid equipment fault prediction and analysis.

In 2022, Sivaiah and Rao [28] addressed the scalability issue of Frequent Pattern Mining in Big Data. They emphasized the need for incremental FIM

algorithms for dynamic databases. Incremental mining uncovers patterns, aiding data analysis and decision-making in various domains. This paper provided a comprehensive survey of incremental pattern mining and its methodologies.

In 2022, Pazdor [29] introduced a method for discovering quantitative frequent patterns in big data mining, expanding beyond traditional Boolean itemset mining. Their Q-Eclat algorithm, a modification of Eclat, excelled in mining quantitative patterns, surpassing the existing MQA-M algorithm's speed for quantitative horizontal frequent pattern mining.

In 2022, Li [30] reformulated frequent and high utility itemsets mining (FHUIM) as a bi-objective problem solvable by multi-objective evolutionary algorithms (MOEAs). To enhance MOEA efficiency for larger databases, FHUIM-BOEA-IMS was introduced. This approach employs an improved mutation strategy favoring items with higher support and utility, leading to better FHUI discovery. Results on four datasets demonstrated FHUIM-BOEA-IMS's superior convergence and final solutions compared to baseline methods.
In 2022, Meruva and Venkateswarlu [31] recognized the significance of identifying relationships between items in common item-set discovery. They highlighted the limitations of the support-confidence framework used in traditional association mining, which fails to consider utility elements. To address this, they proposed a utility-based mining method using a tree structure approach. Experimental results on benchmark datasets demonstrated the method's effectiveness in applications like e-commerce, web mining, and healthcare.

In 2023, Qian et al. [32] introduced rare itemset mining in uncertain databases, presenting two algorithms for probabilistic rare itemset discovery and pruning strategies to reduce computational costs, confirming their effectiveness through experiments. This pioneering work fills a gap in the study of rare itemsets.

In 2023, Siva and Chaudhari [33] emphasized the importance of pattern mining for real-time decision-making. They introduced a lightweight model, CSLFC-HUIM, for robust high-utility itemset mining. This model uses an improved cumulative summary list structure (CULS) and lift-score as a minimum threshold, reducing database scanning and execution time, making it suitable for various pattern mining and business intelligence tasks.

In 2023, Wu et al. [34] addressed the need to extract valuable insights from big data. They introduced a method for mining skyline frequent utility patterns, overcoming limitations of traditional high utility and frequent itemset mining. Their parallel Spark-based algorithm efficiently identified skyline points in large and multi-dimensional datasets, offering effective mining of skyline frequent utility itemsets, as demonstrated in experimental data analysis.

In 2023, Tian et al. [35] tackled the challenges of uncertain frequent pattern mining by introducing a novel approach. They used multiple expected minimum support measures to enhance item value discrimination and address the limitations of a single probabilistic threshold. Their algorithm, incorporating ensembled conditional item-wise supports and pruning strategies, significantly improved the information precision of uncertain frequent itemset mining, as confirmed in extensive experiments.

In 2023, Zhao et al. [36] addressed the need for advanced data mining techniques in the era of big data. They introduced a method for mining closed itemsets and generators in uncertain data, significantly reducing the number of itemsets while retaining crucial information. The algorithm, validated on various datasets, proved both correctness and efficiency.

The reviewed papers from 2022 and 2023 represent significant advancements in data mining. They cover a wide range of applications, including incremental methods for dynamic databases, enhanced fault prediction in power grids, and the scalability of frequent pattern mining. Notable contributions include the development of innovative algorithms for quantitative pattern mining, utility-based approaches for itemset discovery, and efficient solutions for mining rare itemsets in uncertain databases. These papers introduce novel techniques for skyline pattern mining and enhance information precision in uncertain frequent itemset mining. They collectively enrich the data mining field and offer valuable solutions for a diverse set of challenges.

3.Discussion and analysis

The studies conducted in 2022 and 2023 have made significant strides in the field of data mining, offering innovative solutions to various challenges. These

studies have introduced novel methodologies, which have the potential to advance data mining techniques across different domains.

First and foremost, the emphasis on dynamic database updates is crucial in the era of rapidly evolving data. The incremental method introduced in 2022 by Hong et al. [26] represents a promising approach. It overcomes the limitations of batch-based methods, ensuring that the knowledge remains correct and up-to-date. However, the challenge lies in maintaining data accuracy in the face of continuously incoming data. Rapid data influx may introduce complexities that need to be carefully addressed to ensure the reliability of the method.

Efficient fault prediction in power grids, as demonstrated by He and Lu in 2022, [27] is another commendable achievement. While the proposed model outperforms other algorithms in terms of training time and accuracy, it must be adaptable to diverse grid structures and real-world complexities. Power grids come in various forms and sizes, and ensuring the model's efficacy across different scenarios is a challenge that must be tackled.

Scalability is a recurring theme in data mining, as highlighted by Sivaiah and Rao's 2022 [28] study. Their focus on incremental mining methods for Big Data is commendable, but the scalability of such methods, particularly on exceptionally large datasets with varying data types, remains a significant challenge. Data mining in the era of Big Data necessitates efficient and effective techniques that can handle the complexities of vast and diverse data sources.

Quantitative pattern mining, as extended by Pazdor [29] in 2022, offers a promising approach to discover quantitative patterns in large datasets. However, the scalability of this method on exceptionally large datasets could be a limitation. The efficiency and effectiveness of the algorithm should be evaluated in the context of vast and complex data sources.

Multi-objective evolutionary algorithms (MOEAs), as introduced by Li in 2022 [30], present an exciting avenue for data mining. While the approach demonstrated superior convergence and final solutions, its adaptability to different databases and its performance on exceptionally large datasets should be explored further.

Utility-based mining, emphasized by Meruva and Venkateswarlu in 2022 [31], represents an important step in the direction of considering utility elements in association mining. However, adapting this approach to various domains and data structures may present challenges.

Mining rare itemsets in uncertain databases, as initiated by Qian et al. [32] in 2023, fills a gap in the study of rare itemsets. Nevertheless, the algorithm's performance on complex and highly uncertain datasets warrants further investigation.

Real-time decision-making, a crucial aspect in data-driven applications, was addressed by Siva and Chaudhari [33] in 2023. Their lightweight model for high-utility itemset mining is promising but must be tested in various real-time decision-making scenarios.

Skyline pattern mining, as presented by Wu et al. in 2023 [34], offers an effective solution for extracting valuable insights from big data. However, identifying skyline points in complex and multi-dimensional datasets efficiently remains a challenge.

Uncertain frequent itemset mining, a novel approach introduced by Tian et al. in 2023 [35], enhances information precision. Its performance on highly uncertain datasets and its adaptability to different data scenarios should be examined.

Lastly, the study by Zhao et al. [36] in 2023 focused on mining closed itemsets in uncertain data. While it significantly reduces the number of itemsets while retaining critical information, its applicability to various types of uncertain data and its performance on complex datasets are areas that require further investigation.

These studies have introduced valuable contributions to data mining, offering innovative solutions to various challenges. However, addressing the scalability, adaptability, and performance limitations, especially in the context of complex and diverse datasets, is essential for the practical application of these methodologies in real-world data mining scenarios. Future research should focus on refining these approaches to meet the evolving demands of data mining in the digital age.

4.Conclusion

The reviewed studies from 2022 and 2023 represent significant strides in the field of data mining. Researchers introduced innovative solutions to

various data mining challenges. From dynamic database updates to efficient fault prediction and scalable frequent pattern mining, these studies enhance data mining techniques for modern data analytics. The presented methods for quantitative pattern mining, utility-based approaches, mining rare itemsets, real-time decision-making, and uncertain frequent itemset mining offer potential solutions for practical applications. However, it is crucial to address challenges related to scalability, adaptability to diverse data types, and performance on complex datasets. Future research must focus on refining these approaches to meet the evolving demands of data mining in the digital age.

Acknowledgment
None.

Conflicts of interest
The authors have no conflicts of interest to declare.

References
[1] Li J, Gan W, Gui Y, Wu Y, Yu PS. Frequent itemset mining with local differential privacy. In proceedings of the 31st ACM international conference on information & knowledge management 2022 (pp. 1146-55).
[2] Wu JM, Teng Q, Huda S, Chen YC, Chen CM. A privacy frequent itemsets mining framework for collaboration in iot using federated learning. ACM Transactions on Sensor Networks. 2023; 19(2):1-5.
[3] Alsaeedi HA, Alhegami AS. An incremental interesting maximal frequent itemset mining based on FP-growth algorithm. Complexity. 2022; 2022.
[4] Kumar M, Dubey AK. An analysis and literature review of algorithms for frequent itemset mining. International Journal of Advanced Computer Research. 2023; 13(62):1-7.
[5] Cleland ZW, Dao KA, Dao TH. Detecting changes in spatial characteristics of Colorado human-caused wildfires using APRIORI-based frequent itemset mining. Computers, Environment and Urban Systems. 2023; 101:101941.
[6] Dubey AK, Shandilya SK. Exploiting need of data mining services in mobile computing environments. In international conference on computational intelligence and communication networks 2010 (pp. 409-14). IEEE.
[7] Akash MB, Mandal I, Al Mamun MS. Backward support computation method for positive and negative frequent itemset mining. Journal of Data Analysis and Information Processing. 2023; 11(1):37-48.
[8] Chen Y, Gan W, Wu Y, Philip SY. Privacy-preserving federated mining of frequent itemsets. Information Sciences. 2023; 625:504-20.
[9] Raj S, Ramesh D. PartEclat: an improved Eclat-based frequent itemset mining algorithm on spark clusters using partition technique. Cluster Computing. 2022; 25(6):4463-80.
[10] Jamsheela O, Raju G. SR-mine: Adaptive transaction compression method for frequent itemsets mining. Arabian Journal for Science and Engineering. 2022; 47(8):9641-57.
[11] Magdy M, Ghaleb FF, Mohamed DA, Zakaria W. CC-IFIM: an efficient approach for incremental frequent itemset mining based on closed candidates. The Journal of Supercomputing. 2023; 79(7):7877-99.
[12] Zhao Z, Lan L, Wang B, Lai J. Verifiable privacy-preserving outsourced frequent itemset mining on vertically partitioned databases. Electronics. 2023; 12(8):1952.
[13] Lu J, Xu W, Zhou K, Guo Z. Frequent Itemset mining algorithm based on linear table. Journal of Database Management (JDM). 2023; 34(1):1-21.
[14] Al-Ghanimi OA, Khafaji HK. A logic design-based approach for frequent itemsets mining using LCO algorithm. International Journal of Intelligent Engineering & Systems. 2023; 16(2):591-606.
[15] Lakshmi N, Krishnamurthy M. Association rule mining based fuzzy manta ray foraging optimization algorithm for frequent itemset generation from social media. Concurrency and Computation: Practice and Experience. 2022; 34(10):e6790.
[16] Faisal M, Rahman TK. Determining rural development priorities using a hybrid clustering approach: a case study of South Sulawesi, Indonesia. International Journal of Advanced Technology and Engineering Exploration. 2023; 10(103):696-719.
[17] Dubey AK, Dubey AK, Agarwal V, Khandagre Y. Knowledge discovery with a subset-superset approach for Mining heterogeneous data with dynamic support. In CSI sixth international conference on software engineering (CONSEG) 2012 (pp. 1-6). IEEE.
[18] Dubey AK, Shandilya SK. A novel J2ME service for mining incremental patterns in mobile computing. In information and communication technologies: international conference, iCT 2010, Kochi, Kerala, India. Proceedings 2010 (pp. 157-64). Springer Berlin Heidelberg.
[19] Vats S, Sunny S, Sagar BB. Apriori-based algorithms with a decentralized approach for mining frequent Itemsets: a review. Advanced Computer Science Applications. 2023:145-55.
[20] Seddiki I, Nouioua F, Barkat A. Extracting sequential frequent itemsets from probabilistic sequences database. International Journal of Information Technology. 2023:1-7.
[21] Ananthi J, Sengottaiyan N, Anbukaruppusamy S, Upreti K, Dubey AK. Forest fire prediction using IoT and deep learning. International Journal of Advanced Technology and Engineering Exploration. 2022; 9(87):246-56.
[22] Dhanaseelan R, Jeyasutha M. A novel fuzzy frequent Itemsets mining approach for the detection of breast cancer. In research anthology on medical informatics in breast and cervical cancer 2023 (pp. 511-31). IGI Global.

[23] Qian K, Gao S, Yu L. Marginal frequent itemset mining for fault prevention of railway overhead contact system. ISA Transactions. 2022; 126:276-87.
[24] Mushi PK, Ngondya D. Prediction of mathematics performance using educational data mining techniques. International Journal of Advanced Computer Research. 2021; 11(56):83-102.
[25] Cui Y, Gan W, Lin H, Zheng W. FRI-miner: fuzzy rare Itemset mining. Applied Intelligence. 2022:1-6.
[26] Hong TP, Hung WT, Huang WM, Tsai YC. Incremental fuzzy utility mining with tree structure. In international conference on big data (Big Data) 2022 (pp. 6202-6). IEEE.
[27] He W, Lu W. Method of equipment fault diagnosis based on frequent Itemsets mining. In international conference on measuring technology and mechatronics automation (ICMTMA) 2022 (pp. 523-7). IEEE.
[28] Sivaiah B, Rao RR. A survey on fast and scalable incremental frequent Item set methods for big data. In international conference on intelligent controller and computing for smart power (ICICCSP) 2022 (pp. 1-5). IEEE.
[29] Pazdor AG, Leung CK, Czubryt TJ, Lu J, Popov D, Raval S. Social network analysis of popular YouTube videos via vertical quantitative mining. In IEEE/ACM international conference on advances in social networks analysis and mining (ASONAM) 2022 (pp. 303-7). IEEE.
[30] Li C. Frequent and high utility Itemsets mining based on bi-objective evolutionary algorithm with an improved mutation strategy. In 21st international symposium on distributed computing and applications for business engineering and science (DCABES) 2022 (pp. 15-8). IEEE.
[31] Meruva SR, Venkateswarlu B. A fast and effective tree-based mining technique for extraction of high utility Itemsets. In 6th international conference on electronics, communication and aerospace technology 2022 (pp. 1393-9). IEEE.
[32] Qian W, Li H, Huang H, Yuan M, Xu Y, Sun G. Fast rare itemset mining in uncertain database. In 3rd international conference on neural networks, information and communication engineering (NNICE) 2023 (pp. 613-20). IEEE.
[33] Siva S, Chaudhari S. Cumulative summary list driven lightweight frequent closed high utility Itemset mining. In 2nd international conference for innovation in technology (INOCON) 2023 (pp. 1-6). IEEE.
[34] Wu JM, Zhou H, Lin JC, Wang K, Liu S, Li R. A novel spark-based algorithm for mining frequent utility patterns. In 6th international conference on artificial intelligence and big data (ICAIBD) 2023 (pp. 99-104). IEEE.
[35] Tian W, Li F, Liu Y, Wang Z, Zhang T. Depth-first uncertain frequent itemsets mining based on ensembled conditional item-wise supports. In international conference on intelligent supercomputing and BioPharma (ISBP) 2023 (pp. 121-8). IEEE.
[36] Zhao L, Chen C, Tian W. Mining frequent closed Itemsets and generators over uncertain data. In 6th international conference on electronic information and communication technology (ICEICT) 2023 (pp. 453-8). IEEE.

Abhilash Behera is pursuing M.Tech in Computer Science & Engineering at Madhyanchal Professional University, Bhopal (MP) , India, and holds a B.Tech degree in Computer Science & Engineering from Biju Patnaik University Of Technology, Odisha, Sundargarh (SEC),India. His primary area of interest is Data Mining.
Email: abhilashbehera7448@gmail.com