10.48047/jocaaa.2024.33.08.239

High Utility Itemset Mining: A Thorough Examination of Data Mining Methods, Algorithms, and Uses

Dharmbir[1]*, Ram Kinkar Pandey[2], Rajiv Kumar[3]
Department of Computer Science & Engineering, Arni University, Himachal Pradesh[1,2]
School of Computer Science Engineering and Technology, Bennett University, Greater Noida[3]
[1*]dharmvryadav@gmail.com, [2]Dr.ramkpandey@gmail.com, [3]rajiv.kumar@bennett.edu.in

Abstract

The High Utility Itemset Mining (HUIM) is an important revolution in the data mining world as it provides organizations and researchers with the option of mining patterns that are not only common but also profitable and relevant in the context. A distinction between the HUIM and traditional association rule mining is that utility measures are used to integrate the measures like support and confidence, unlike the traditional association rule mining which mainly uses these measures. The definitions of utility can vary according to the field of application: the profit margins in the retail business, the effectiveness of treatment in medical care, or the risk level in finance and fraud detection.

The discipline of HUIM has experienced an impressive development within the last 20 years. The first algorithms like Two-Phase method laid the ground work by introducing candidate generation with utility upper bound and the later innovations like UP-Growth and HUI-Miner offered more efficient data structures like utility lists and utility pattern tree. Subsequent methods such as FHM and EFIM further improved the pruning algorithms and memory management to allow HUIM to run on huge datasets. In addition to the development of algorithms, HUIM has been used in other areas including e-commerce recommendation, personalized marketing, medical decision support, financial fraud detection, and optimization of the supply chain.

Nevertheless, despite the achieved success, HUIM is still experiencing the urgent issues such as computational complexity, scalability to big data platforms, work with dynamic and streaming data, and sensitivity of privacy when using utility-oriented mining. Future research directions involve incorporating HUIM into sophisticated machine learning, reinforcement learning, blockchain and big data structures to ensure HUIM is flexible, and secure and expandable in real time environments.

In this paper, a total review of HUIM is given, including the conceptual background, algorithmic development, comparative studies, applications, major challenges, and the future research trends. This review will be a useful source of knowledge to researchers, practitioners, and decision-makers interested in leveraging the power of HUIM to extract value-driven insights through synthesis of existing literature and identification of important gaps.

10.48047/jocaaa.2024.33.08.239

1. Introduction

The amount of data created by businesses, governments, and individuals in the contemporary digital age is impressive. With online retail interactions up to the healthcare records and sensor information on the Internet of Things (IoT), organizations are currently confronted with a hallucinating amount of structured and unstructured data. Due to this flood of information, it has become a challenge and an opportunity to extract meaningful knowledge[1]. The field of data mining has occupied the center stage in resolving this issue, with methods that include integration of classification and clustering, frequent pattern and association rule mining.

These include Frequent Itemset Mining (FIM) and Association Rule Mining(ARM) whose initial popularity was due to their capability to detect associations of co-occurring items within large transaction databases. The market basket analysis is based on seminal algorithms like Apriori (Agrawal and Srikant, 1994) and its further developments (FP-Growth, Han et al., 2000), which helped the retailer to discover relationships such as those in the case of bread and butter, where the latter is also purchased by customers who buy bread[2]. Such insights were of use but their weakness was soon revealed. Patterns that were high in frequency did not necessarily mean high value particularly in a commercial environment where the bottom line was profitability.

This gap was a reason that High Utility Itemset Mining (HUIM) was developed that extends beyond frequency to finding itemsets that optimize utility. Utility in this sense can be defined as quantitative importance of an itemset, which may be in terms of profit, revenue, saving of cost, risk, or even user-specified importance. As an example, in a supermarket dataset, bread and milk may be common, but things such as wine and fancy cheese, though bought less, may drive much more earnings[3].

HUIM has its importance in its correspondence to the real world business and decision making needs. Compared to ARM, which in most cases overwhelms analysts with the excessive amount of rules, many of which are empty, HUIM concentrates on itemsets that have real significance in terms of utility. This valuable especially in areas like:

- E-commerce, in which profitability and personal recommendations are the sources of competitiveness.
- Healthcare where high-utility treatment combinations can be used to optimize patient outcomes.
- Fraud detection, in which anomalies are infrequent, yet high valuable.

10.48047/jocaaa.2024.33.08.239

- The supply chain management, where the inventory optimization on the high-utility patterns can lead to the cost reduction and enhancement of the efficiency.

This paper will set out to review HUIM comprehensively in terms of its theoretical basis, the development of the algorithm, challenges and applications of the algorithm. It also addresses future research directions, especially the combination of HUIM with machine learning and reinforcement learning and blockchain, which are likely to increase its flexibility and effectiveness in dynamic settings.

2. Background and Motivation

2.1 The development of Frequent/HUIM The History of Frequent to HUIM.

The history of HUIM cannot be separated to the history of frequent itemset mining. Systematic mining of frequent patterns First of its kind, the Apriori algorithm (1994) used candidate generation and pruning methods based on minimum support values[4]. Although widely used, Apriori had the downside of the combinatorial explosion problem, since its candidate itemsets increased exponentially with the size of the dataset.

Later development such as an FP-Growth algorithm (2000) was more efficient, building a smaller-sized data structure called the Frequent Pattern Tree (FP-Tree) that did not require candidate generation. Equally, Eclat algorithm proposed a vertical database with transaction ID (TID) sets to mine frequent itemsets faster[5]. Although these approaches had a great development on pattern mining, their use of support and confidence as measures of evaluation restricted their usefulness in practice.

Take an example: in a supermarket dataset, there may be a bread, milk, and sugar items which appear together quite frequently but may bring very small profit. On the other hand, luxury goods like electronics, jewelry or expensive wine may seem less frequent, but they will give out disproportionately higher returns[6]. In the traditional frequent itemset mining, the frequent low-value itemsets would be given priority whereas high-value infrequent sets would be ignored.

In order to overcome this drawback, scientists developed the utility-based mining idea, according to which the relevance of objects is determined by such measures as profit, cost, risk, or importance, not by the frequency of occurrence[7]. This change of paradigm resulted in the development of High Utility Itemset Mining (HUIM) in the early 2000s.

10.48047/jocaaa.2024.33.08.239

2.2 Motivation for HUIM

HUIM is driven by the fact that it produces actionable value-based insights:

- Business Profitability: Retailers and online market places must find the right products that sell as well as those that yield high profitability. HUIM enables this through the conglomeration of frequency and profit analysis.
- Healthcare Effectiveness: Within the healthcare environment, HUIM can demonstrate medicine combinations or treatment regimens that have the greatest therapeutic benefit to cost.
- Fraud and Risk Management: Fraudulent transactions are not that common but are usually high-value transactions. HUIM facilitates the early identification of such anomalies.
- Scalability in Big Data: Dynamic e-commerce and IoT are increasing the size of big data, so HUIM algorithms must be scalable to terabytes of transactional data.

HUIM, therefore, enables a framework that converts data mining outputs into strategic organizational objectives, and this is invaluable in the age of data-driven decision-making[8].

3. Problem Statement

The main issue that HUIM tries to solve is based on the weaknesses of the traditional frequent itemset mining. In particular, frequent itemset mining only gives credence to patterns in terms of support (frequency) and confidence (conditional probability). Although these measures are useful in determining co-occurring items, they do not reflect on whether the patterns discovered are of benefit in practice.

For instance:

- A retail data set can show that pens and erasers are often sold at the same time, yet the margin of this bundle is low[9].
- On the other hand, laptops and prolonged warranties are not bought regularly but have high profit margins.

The classical frequent itemset mining would focus on the former trend and overlook the latter causing poor business decisions.

The definition of the problem of HUIM can be thus formulated:

10.48047/jocaaa.2024.33.08.239

- High-Utility Itemset Discovery: Find itemsets in a transactional database with aggregate utility that is larger than a user-specified threshold. Utility can be calculated depending on the item profit, quantity or weight.
- Scalability and Efficiency: Produce design algorithms that can handle the exponential search space with limited computational overhead.
- Dynamic Utility Adaptation: Model real-world environments where item utility (e.g. profit margins, treatment costs, risk value) is a time-varying parameter.
- Assessment of Usefulness: Work out useful utility scales that are both comprehensive and practical.

This is because there is a trade-off between the computational feasibility and the practicality of the HUIM algorithms in that they need to operate effectively on a large-scale data but the resultant output needs to reflect back on the organizational strategies.

4. Literature Review

4.1 Classical Data Mining Methods.

HUIM starts with the classical frequent itemset mining. Some of the most important approaches are:

- Apriori (1994): According to the principle that every subsets of frequent itemset should also be frequent. It is not efficient with large datasets because it depends on candidate generation[11].
- FP-Growth (2000): Constructs a small FP-Tree which represents itemsets and mines patterns without the need to produce candidates. Effective, but restricted to frequency based mining.
- Eclat (2000): It is a vertical database format that uses transaction IDs as an aid in faster intersection-based mining. More effective when data are sparse.

Although these algorithms increased the pattern discovery, their failure to integrate utility led to the introduction of HUIM.

4.2 HUIM Algorithms Evolution.

Two-Phase Algorithm (2005)

The initial important HUIM method was the Two-Phase algorithm proposed by Liu et al. It produces a huge quantity of candidate itemsets and narrows them down by utility upper bounds

10.48047/jocaaa.2024.33.08.239

like Transaction Weighted Utility (TWU). Although fundamental, it is computationally inefficient and consumes much memory on large data sets.

UP-Growth (2010)

UP-Growth proposed Utility Pattern Tree (UP-Tree) that stores the information about the transactions in a compressed format[12]. Pruning techniques applied in the algorithm include Discarding Global Node Utility (DGN) and Discarding Local Node Utility (DLN) in order to cut down on candidate itemsets.

HUI-Miner (2012)

HUI-Miner substituted candidate generation with utility-list format, whereby an itemset is represented by a list of its transaction ID and utility data. This greatly lowers memory overhead and efficiency.

FHM (2014)

Fast High-Utility Miner (FHM) algorithm is based on HUI-Miner, with the addition of Estimated Utility Co-occurrence Pruning (EUCP) pruning unpromising itemsets early in the mining process. FHM also provides considerable enhancements in run time performance.

EFIM (2015)

One of the fastest HUIM algorithms is now assumed to be the Efficient High-Utility Miner (EFIM). EFIM applies a depth-first search, sophisticated pruning techniques, and compact data structures to search HUIs in very large datasets.

Hybrid and Deep Learning Approaches (2020 and on)

Recent studies combine HUIM and machine learning and deep learning methods and combine HUIM interpretability with prediction. They may be HUIM based on reinforcement learning, and deep neural network-based hybrid approaches to adaptive pattern recognition[16].

5. Applications of HUIM

High Utility Itemset Mining has a broad range of applications in the areas where data-driven decision-making is the most crucial element. HUIM has been found to be more in tune with the real world needs by the inclusion of profitability, cost, or other contextual utility metrics than

10.48047/jocaaa.2024.33.08.239

traditional frequent itemset mining. Some of the most noticeable application areas are listed below:

Figure 1. Machine Learning Applications In Different Domains.

5.1 Retail and E-commerce

The most immediate beneficiaries of HUIM are still retail and e-commerce. The common co-purchase patterns are identified through traditional market basket analysis, however HUIM takes

10.48047/jocaaa.2024.33.08.239

this a step further by incorporating profit margins, promotion costs and seasonal trends in the analysis. As the example of bread and milk, which could be a regular purchase, may not result in high profit[19]. Conversely, wine and imported cheese combination, although less common, have a lot more utility.

Applications include:

- Product bundling: Formulation of profitable item bundling according to the high-utility co-purchase behavior.
- Dynamic pricing: The approach involves determining high-utility items to maximize the discount strategy without compromising profitability.
- Cross-selling and recommendations: Recommendation of complementary goods which will maximise revenue[21].
- Inventory optimization: Optimizing stock levels to the high-profit products and not to high-frequency but low-profit products.

Massive competitors like Amazon, Flipkart, and Alibaba use these techniques to enhance their recommendation engines and revenue maximization plans.

5.2 Healthcare and Bioinformatics.

Patient records, treatment plans, and diagnostic outcomes are some of the common healthcare data, and HUIM has rich grounds to thrive. In healthcare, compared to retail where profit is the measure of utility, it may take on a different meaning to include, but not be limited to, patient recovery rate, treatment cost-effectiveness, or quality-adjusted life years (QALY).

Applications include:

- Drug discovery: To find drug combinations that are maximally therapeutic.
- Treatment optimization: Identifying high-utility treatment pathways to particular conditions, at a cost/effectiveness trade-off.
- Genomic studies: mining of DNA sequence patterns in which utility is related to biological significance or impact of mutation.
- Patient management: Allocation of resources in hospitals according to high-utility patterns of patient-care.

HUIM is especially applicable to personalized medicine, where therapies are developed to optimize the effect on individual patients and not on generic populations.

10.48047/jocaaa.2024.33.08.239

5.3 Finance and Fraud Detection

Financial sector is very utility-oriented and HUIM is an ideal fit. In this case the utility tends to be the monetary value, risk exposure or loss. One of the most sensitive applications is fraud detection.

- Credit card fraud identification: HUIM can detect high value, but rare fraudulent transactions because it is impact-based rather than frequency-based[22].
- Stock market analysis: Determination of the high utility trading patterns to optimize the portfolio.
- Risk management: Mining transaction patterns that expose to high-risk clients or sectors.
- Insurance fraud detection: Finding anomalous, but valuable, claims which are missed by frequency-based techniques.

HUIM is also more beneficial than traditional anomaly detection since it shows the rare but high-impact cases which directly impact profitability.

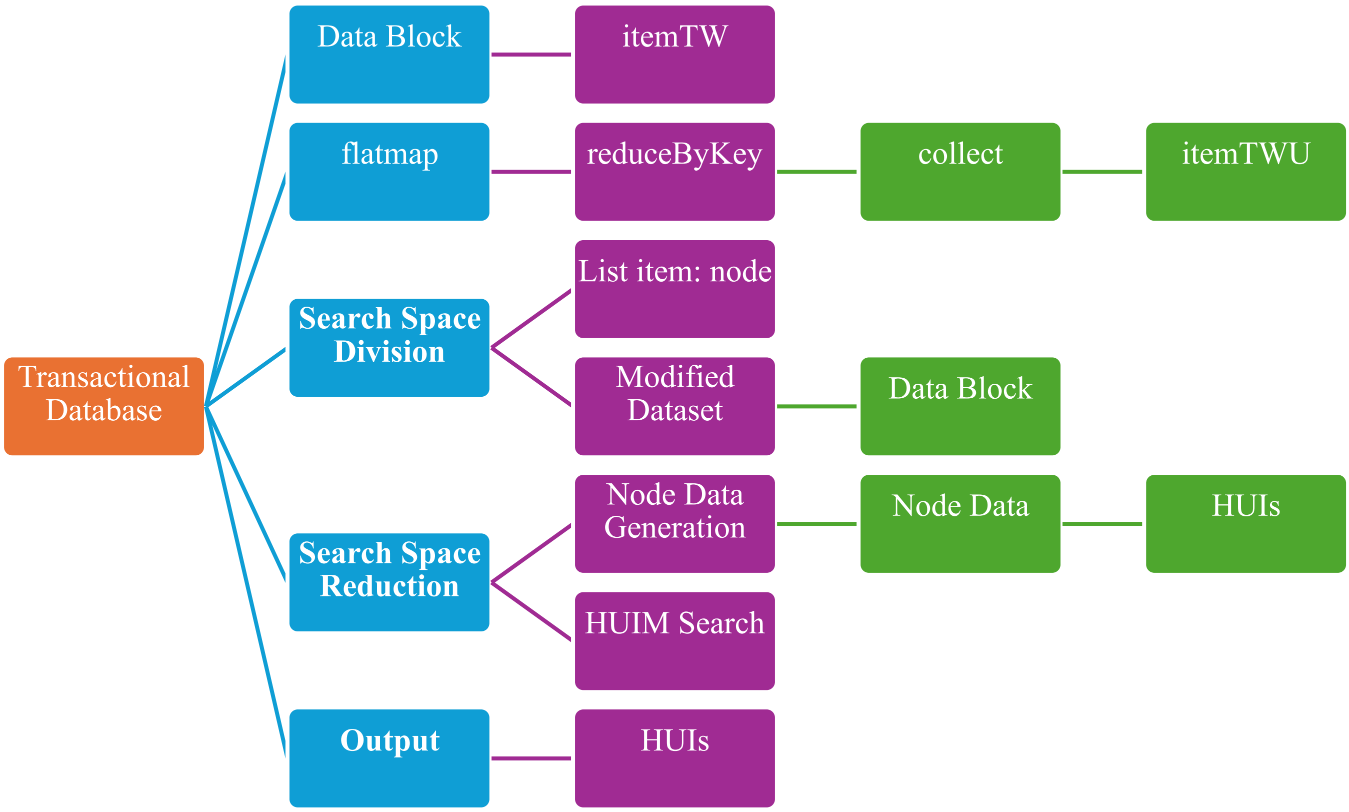

Figure 2. EAHUIM Flow Graph.

10.48047/jocaaa.2024.33.08.239

5.4 Supply Chain and Logistics

HUIM assists firms in having optimal procurement, warehousing, and distribution plans in the supply chain management by focusing more on cost and efficiency as opposed to the frequency of the raw materials.

Applications include:

- Vendor analysis: Recognizing suppliers that make high-utility contribution to efficiency and profitability.
- Inventory planning: It is more prudent to focus storage and transportation on high utility items.
- Route optimization: Minimizing transportation strategies according to mining logistics data to identify high-utility transportation strategies.

In the case of global supply chains, HUIM can help decision-makers to put resources in areas where they can give the highest returns.

5.5 Other Domains

- Education: Determining learning materials or instructional methods that produce high performance results by students[23].
- Cybersecurity: Learning attack patterns that have high utility to focus on defense.
- IoT and Smart Cities: Optimizing energy use by digging into device usage patterns where utility is quantified in terms of energy saved.
- Telecommunications: Determining lucrative calling or data plans that will give the maximum revenue per customer.

HUIM has therefore developed into an interdisciplinary instrument where it is used in commerce, to community health and infrastructure control.

6. Methodological Directions

HUIM methodologies are not fixed; they have been modified to meet the needs of challenges in scaling, optimization of memory and dynamic utility definitions. In this section, an overview of methodological innovations is given.

6.1 Data Structures in HUIM

Effective data format is important in minimizing computational work:

10.48047/jocaaa.2024.33.08.239

- Utility Pattern Tree (UP-Tree): Pruning is supported and transactions are compressed.
- Utility List: This is the itemsets and their transaction IDs and utility, without candidate generation.
- Vertical Layout Databases: Allow itemset mining intersection operations to run more quickly.

These data structures are trade off between memory usage and performance, and HUIM can be scaled to bigger datasets.

6.2 Pruning Strategies

The exponential search space needs to be controlled by pruning:

- Transaction Weighted Utility (TWU): This is a measure of an upper bound to eliminate unpromising itemsets at an early stage.
- Discarding Global/Local Node Utility (DGN/DLN): This is applied in the UP-Growth to minimize overestimation of utility.
- Estimated Utility Co-occurrence Pruning (EUCP): Less utility list joins are made by removing non-promising candidates.
- Local Utility Thresholds: Dynamically tune pruning, according to dataset properties.

Pruning techniques are very effective in boosting performance by computing only on patterns that are promising.

6.3. The parallel and distributed HUIM

Parallel HUIM algorithms have been designed with the emergence of big data:

- HUIM based on MapReduce: Scatters data across cluster to do large-scale mining.
- GPU-accelerated HUIM: Takes advantage of massive parallelism to compute fast.
- Cloud-native HUIM: Interoperates with cloud architecture real-time analytics.

These approaches will make HUIM sustainable in the age of the scale of distributed computing.

6.4 Incremental and Stream Mining.

Currently, with the spread of real-time applications, any use of static datasets is becoming rarer:

- Incremental HUIM: Re-mining of the complete dataset is not done because results are updated after new transactions are added[24].
- Stream HUIM: High-speed data streams, based on a window model or a landmark model.

10.48047/jocaaa.2024.33.08.239

This flexibility is crucial in areas like finance, IoT and cybersecurity, as information comes in at a frequent rate.

7. Comparative Analysis

To understand the performance of HUIM algorithms, researchers often compare them along three dimensions: **efficiency**, **memory consumption**, and **scalability**.

Algorithm	**Approach**	**Strengths**	**Weaknesses**
Two-Phase	Candidate generation	Simple, foundational	Poor scalability
UP-Growth	UP-Tree + pruning	Reduces candidates	Complex tree management
HUI-Miner	Utility list	Avoids candidates	Memory intensive
FHM	Utility list + EUCP	Efficient pruning	Still limited for very large data
EFIM	Depth-first + pruning	Extremely fast	Complexity in implementation

A comparison (Fournier-Viger et al., 2019) of EFIM and FHM with previous algorithms is always presented with a superiority in runtime and memory consumption. Nevertheless, actual performance is a function of data set size, sparsity, and domain utility definitions.

10.48047/jocaaa.2024.33.08.239

8. Future Directions

The future of HUIM is in incorporating it with state-of-the-art computational paradigm and dealing with the existing shortcomings.

8.1. Machine Learning Interaction.

- Hybrid HUIM + Deep Learning: A neural network-based utility-based pruning guidance.
- Reinforcement Learning: Adaptive HUIM which learns to use the best utility thresholds dynamically[25].
- Explainable AI: It is necessary to ensure that HUIM results are understandable to decision-makers.

8.2 Real-time and Streaming Applications.

- Low-latency HUIM development to detect fraud and IoT.
- Dynamic utility evaluation adaptive thresholds.

8.3 Privacy-preserving HUIM

As the problem of data privacy raises more and more concerns, privacy-preserving HUIM seeks to secure confidential data by:

- Privacy differentiation.
- Federated HUIM in which data is mined without centralizing sensitive data.

8.4 Integration with Blockchain and Big Data.

- Finance and healthcare Finance and healthcare Blockchain can make HUIM transactions secure and auditable.
- HUIM can now be scaled to petabytes using Big Data platforms like Spark and Flink.

9. Conclusion

High Utility Itemset Mining has become a strong development of the classical frequent itemset mining shifting the emphasis towards utility. This renders it especially applicable to real world domains where profitability, cost, risk, or impact is more important than raw frequency.

HUIM has evolved in more than 20 years since its first candidate-based methods such as Two-Phase to high-quality algorithms such as EFIM and FHM that use complex data structures and pruning techniques. It has been used in a wide variety of applications in retail, healthcare, finance, supply chain, cybersecurity, and more, and has demonstrated its flexibility and usefulness.

10.48047/jocaaa.2024.33.08.239

Nevertheless, HUIM still has some issues, including scalability in big data operations, utility adaptation by dynamic needs, and privacy. The improvement of HUIM in the future will be based on the incorporation of deep learning, reinforcement learning, blockchain, and the big data frameworks.

HUIM helps data analytics to keep up with the demands of organizations, industries and societies by closing the gap between theoretical data mining and practical utility-driven results. It is not only a computational challenge but also a pillar of the data-driven decision-making in the digital era.

References

1. Abhinav, Krishnamurthy, N., & Jain, R. (2018). Wear Studies of Al2O3-ZrO2·5CaO Composite Coatings for Tribological Applications. International Journal of Engineering & Technology, 7(3.4), 73. https://doi.org/10.14419/ijet.v7i3.4.16750
2. Agrawal, R., & Srikant, R. (1994). Fast algorithms for mining association rules. Proceedings of the 20th International Conference on Very Large Data Bases (VLDB), 487–499.
3. 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, 1–12.
4. Liu, Y., Liao, W. K., & Choudhary, A. (2005). A two-phase algorithm for fast discovery of high utility itemsets. Advances in Knowledge Discovery and Data Mining, 689–695. Springer.
5. Tseng, V. S., Wu, C. W., Fournier-Viger, P., & Yu, P. S. (2013). Efficient algorithms for mining high utility itemsets from transactional databases. IEEE Transactions on Knowledge and Data Engineering, 25(8), 1772–1786.
6. Liu, Y., Qu, Q., Han, J., & Hu, H. (2012). Mining high utility itemsets without candidate generation. Proceedings of the 21st ACM International Conference on Information and Knowledge Management (CIKM), 55–64.
7. Fournier-Viger, P., Lin, J. C. W., Kiran, R. U., Koh, J. M., & Thomas, R. (2019). A survey of high-utility itemset mining. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 9(4), e1297.
8. Zida, S., Fournier-Viger, P., Lin, J. C. W., Wu, C. W., & Tseng, V. S. (2015). EFIM: A highly efficient algorithm for high-utility itemset mining. Proceedings of the 24th ACM International Conference on Information and Knowledge Management (CIKM), 579–588.

10.48047/jocaaa.2024.33.08.239

9. Deng, Z. H. (2012). PrePost: An efficient N-lists-based algorithm for mining frequent itemsets via prefix-preserving closure extension. Applied Intelligence, 41(1), 69–83.
10. Lin, J. C. W., Fournier-Viger, P., Hong, T. P., & Voznak, M. (2016). A fast algorithm for mining high-utility itemsets with negative unit profits. Knowledge-Based Systems, 111, 283–298.
11. Shie, B. E., Tseng, V. S., & Yu, P. S. (2013). Online mining of temporal maximal utility itemsets from data streams. Proceedings of the 2013 IEEE International Conference on Data Mining (ICDM), 428–437.
12. Lan, G. C., Hong, T. P., & Tseng, V. S. (2014). An efficient projection-based indexing approach for mining high utility itemsets. Knowledge and Information Systems, 38(1), 85–107.
13. Fournier-Viger, P., Lin, J. C. W., Gan, W., & Tseng, V. S. (2016). PHM: Mining periodic high-utility itemsets. Proceedings of the 2016 International Conference on Advanced Data Mining and Applications (ADMA), 64–77.
14. Ahmed, C. F., Tanbeer, S. K., Jeong, B. S., & Lee, Y. K. (2009). Efficient tree structures for high utility pattern mining in incremental databases. IEEE Transactions on Knowledge and Data Engineering, 21(12), 1708–1721.
15. Fournier-Viger, P., Gomariz, A., Campos, M., & Thomas, R. (2014). Fast vertical mining of sequential patterns using co-occurrence information. Proceedings of the 2014 Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD), 40–52.
16. Gan, W., Lin, J. C. W., Fournier-Viger, P., Chao, H. C., & Yu, P. S. (2019). HUOPM: High-utility occupancy pattern mining. IEEE Transactions on Cybernetics, 50(3), 1195–1208.
17. Krishnamoorthy, S. (2018). Mining high utility itemsets with transaction truncation. Expert Systems with Applications, 91, 336–351.
18. Srikant, R., & Agrawal, R. (1996). Mining quantitative association rules in large relational tables. Proceedings of the 1996 ACM SIGMOD International Conference on Management of Data, 1–12.
19. Yun, U., & Ryu, K. (2012). Efficient mining of maximal frequent itemsets using subsumption checking and pruning. Expert Systems with Applications, 39(5), 5629–5641.
20. Lin, J. C. W., Fournier-Viger, P., Wu, C. W., & Hong, T. P. (2017). Efficient mining of high-utility itemsets with multiple minimum utility thresholds. Knowledge-Based Systems, 113, 100–115.

10.48047/jocaaa.2024.33.08.239

21. Tanbeer, S. K., Ahmed, C. F., Jeong, B. S., & Lee, Y. K. (2009). Discovering periodic-frequent patterns in transactional databases. Proceedings of the 13th Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD), 242–253.

22. Fumarola, F., Lanotte, P. F., Pio, G., & Ceci, M. (2016). CloFAST: Closed sequential pattern mining using sparse and vertical id-lists. Knowledge and Information Systems, 48(2), 429–463.

23. Yao, H., Hamilton, H. J., & Butz, C. J. (2004). A foundational approach to mining itemset utilities from databases. Proceedings of the 2004 SIAM International Conference on Data Mining (SDM), 482–486.

24. Tseng, V. S., Fournier-Viger, P., Lin, J. C. W., & Koh, J. M. (2016). High-utility sequential pattern mining. Proceedings of the 2016 International Conference on Big Data Analytics and Knowledge Discovery (DaWaK), 1–18.

25. Gan, W., Lin, J. C. W., Fournier-Viger, P., & Chao, H. C. (2018). A survey of utility-oriented pattern mining. IEEE Transactions on Knowledge and Data Engineering, 31(12), 2233–2254.

26. Fournier-Viger, P., Lin, J. C. W., Vo, B., & Tseng, V. S. (2021). High-utility itemset mining: State-of-the-art and open challenges. Data Science and Engineering, 6(4), 367–388.