Efficient Visualization of Association Rule Mining Using the Trie of Rules

Mikhail Kudriavtsev[1] [a], Andrew McCarren[2] [b], Hyowon Lee[2] [c] and Marija Bezbradica[3] [d]

[1] *Centre for Research Training in Artificial Intelligence (CRT-AI), Dublin City University, Dublin, Ireland*

[2] *Insight Centre for Data Analytics, Dublin City University, Dublin, Ireland*

[3] *Adapt Research Centre, Dublin City University, Dublin, Ireland*

Keywords: Association Rule Mining, Data Visualization, Trie of Rules, FP-tree, Frequent Pattern Tree, Cognitive Load, Visualization Efficiency, Data Mining Techniques.

Abstract: Association Rule Mining (ARM) is a popular technique in data mining and machine learning for uncovering meaningful relationships within large datasets. However, the extensive number of generated rules presents significant challenges for interpretation and visualization. Effective visualization must not only be clear and informative but also efficient and easy to learn. Existing visualization methods often fall short in these areas. In response, we propose a novel visualization technique called the ”Trie of Rules.” This method adapts the Frequent Pattern Tree (FP-tree) structure to visualize association rules efficiently, capturing extensive information while maintaining clarity. Our approach reveals hidden insights such as clusters and substitute items, and introduces a unique feature for calculating confidence in rules with compound consequents directly from the graph structure. We conducted a comprehensive evaluation using a survey where we measured cognitive load to calculate the efficiency and learnability of our methodology. The results indicate that our method significantly enhances the interpretability and usability of ARM visualizations.

1 INTRODUCTION

Association Rule Mining (ARM) is a popular technique in data mining and machine learning that aims to uncover interesting and meaningful relationships within large datasets (Agrawal et al., 1993). These relationships, expressed as ”association rules,” provide valuable insights for decision-making across various domains, such as market basket analysis, healthcare, and fraud detection (Shaukat Dar et al., 2015). However, ARM can produce a vast number of rules, making it difficult to interpret them effectively. Therefore, effective visualization techniques are crucial to help analysts and domain experts make sense of the discovered rules and extract valuable knowledge.

Current visualization approaches for ARM results struggle with significant limitations when displaying a large number of rules while retaining essential information. Existing solutions often either provide incomplete information, limiting the ability to fully interpret and explore the rules, or produce overly large and cluttered charts that are challenging to navigate (Fister et al., 2023; Jentner et al., 2019; Fernandez-Basso et al., 2019). These limitations result in ineffective information display, hindering the practical utility of ARM in real-world applications where understanding complex patterns quickly and accurately can be essential.

In response to these challenges, we developed a novel visualization technique named the ”Trie of Rules.” Our approach addresses the problem of ineffective information display by capturing a wealth of information and maintaining a manageable size when dealing with large datasets. Additionally, it reveals implicitly hidden insights such as substitute pairs or clusters of rules. The Trie of Rules method is based on an adapted Frequent Pattern Tree (FP-tree) structure, traditionally used to visualize transactions. We propose a novel way to interpret this structure to visualize association rules, making our approach both easy to learn and efficient.

A key aspect of our approach is its efficiency. We designed the Trie of Rules to enable users to complete tasks more quickly and accurately when dealing with complex datasets, while maintaining a learnability level comparable to existing methods.

[a] https://orcid.org/0000-0001-9815-5067
[b] https://orcid.org/0000-0002-7297-0984
[c] https://orcid.org/0000-0003-4395-7702
[d] https://orcid.org/0000-0001-9366-5113

Kudriavtsev, M., McCarren, A., Lee, H. and Bezbradica, M.
Efficient Visualization of Association Rule Mining Using the Trie of Rules.
DOI: 10.5220/0012995500003838

In *Proceedings of the 16th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management (IC3K 2024) - Volume 1: KDIR*, pages 72-80
ISBN: 978-989-758-716-0; ISSN: 2184-3228

The main contributions of this paper are as follows:

- **Development of a Visualization Strategy:** We introduce an efficient visualization technique for ARM results that captures extensive information while remaining easy to interpret.
- **Comparison with Popular Methods:** We compared our method with other popular visualization techniques and demonstrated that it outperforms them in terms of efficiency. This was accomplished via a survey with 34 participants, where we measured efficiency and learnability. Our approach allows users to complete tasks more quickly and accurately, while being as easy to learn as existing methods.
- **Confidence Calculation for Compound Consequents:** The Trie of Rules approach introduces a novel property that significantly enhances further exploration of knowledge and increases speed efficiency when examining the ruleset. This feature allows the calculation of Confidence for rules with compound consequents directly from the graph structure, avoiding additional clutter on the plot and making it easier to read and interpret.

This paper is structured as follows: Section 2 provides background information on ARM and related concepts. Section 3 reviews existing visualization methods and their limitations. Section 4 details our proposed Trie of Rules methodology, including the FP-tree background and the visualization approach. Section 5 describes our evaluation methodology, survey construction, and results. Finally, Section 6 summarizes the contributions and suggests directions for future research.

2 BACKGROUND

Association Rule Mining is a data mining technique that aims to discover interesting relationships and patterns within large datasets (Agrawal et al., 1993). The fundamental concepts of ARM include association rules, ruleset, transactions, frequent set, antecedent and consequent, support, and confidence (Geng and Hamilton, 2006; Wu et al., 2010; Luna et al., 2018).

Transactions refer to the records or instances in a dataset, often representing events or actions. In retail, for example, a transaction might correspond to a customer's purchase, where each item bought constitutes a transaction item.

A **frequent set** is a subset of items that frequently occur together in transactions. The identification of frequent sets is a crucial step in ARM, and it involves finding sets of items whose occurrence surpasses a predefined minimum co-occurrence frequency threshold.

An **association rule** is a relationship or pattern that describes the co-occurrence of items in a dataset. It is typically represented as an implication of the form $A \to B$, where A is the **antecedent** and B is the **consequent**. An example of an association rule could be: *If a customer buys item X, they are likely to buy item Y.*

A **ruleset** is a collection of association rules derived from a dataset. The ruleset provides a comprehensive view of the discovered patterns and relationships within the data. Each rule in the ruleset contributes to the understanding of associations between different items.

Metrics are essential for describing association rules, with support, confidence, and lift being the most popular. However, many other metrics exist as well (Hahsler, 2024). These metrics assess the value of rules in various ways. Crucially, they describe the relationship between the antecedent and the consequent, which means they can only be applied to rules. The exception to this is support, which can also be applied to frequent sequences and is frequently used as a metric for the threshold during the mining process.

3 RELATED WORK

Visualizing ARM results is recognized as a challenging task, as indicated by surveys conducted by (Hahsler and Chelluboina, 2011; Fernandez-Basso et al., 2019; Jentner et al., 2019; Alyobi and Jamjoom, 2020; Menin et al., 2021; Fister et al., 2023). The complexity arises from the need to represent rules visually while considering the multitude of associated metrics and distinguishing between antecedents and consequents, leading to various proposed approaches.

Traditionally, rules are presented as plain tables or text-based methods due to their simplicity and familiarity. However, these methods often fail to effectively convey complex relationships, and there is much room for improvement.

Although various methods exist, they can be classified into three distinct groups: scatter plots, matrix-based methods, and graph-based methods.

The **scatter plot** approach, one of the more basic methods, was introduced by (Jr. et al., 1999). This method employs a two or three-dimensional plot (Ong et al., 2002) to depict rules as dots. Although effective in handling a high number of rules, scatter plots lack insight into the structure of rules, requiring manual examination of the text-based representation of the original dataset.

Matrix-based visualization, as presented by (Hofmann and Buhmann, 2000), places antecedent and consequent sets on axes and displays metric values at their intersections. Despite its efficiency in revealing rule components, it suffers from scalability issues, particularly as the dataset size increases. A more modern implementation is provided by (Varu et al., 2022).

An improvement to the matrix-based approach is the **grouped matrix-based visualization**, as proposed by (Hahsler et al., 2017), which alleviates size concerns by grouping similar rules. However, scalability remains a challenge.

Graph-based visualization, widely employed in ARM (Klemettinen et al., 1994; Rainsford and Roddick, 2000; Buono and Costabile, 2005; Ertek and Demiriz, 2006; Fernandez-Basso et al., 2019; Alyobi and Jamjoom, 2020; Menin et al., 2021), provides a clear representation of rule structures. However, the main problem remains how to show all the items in a rule and distinguish between antecedents and consequents. This problem leads to either excessive size of the plot or low interpretability. Current methods rely on the idea that two types of nodes exist—items and rules. Items that go into (directed edge) the rule are antecedents, and edges that go out of a rule node are consequents.

These three main categories are implemented in popular libraries such as arulesViz for R (Hahsler et al., 2017) and arules for Python (Hahsler, 2023).

In conclusion, existing ARM visualization methods exhibit limitations in terms of scalability, interpretability, and representation of rule structures. The proposed methodology in the next section aims to address these challenges by incorporating FP-tree principles to create a more effective visualization.

4 METHODOLOGY

4.1 FP-tree Background

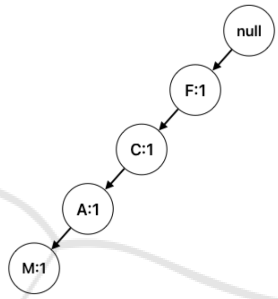
A Frequent Pattern Tree (FP-tree), also known as a **trie** or **prefix tree**, was introduced by (Han et al., 2004). It is commonly used in the rule mining process and is known for its efficiency (Bodon and Rónyai, 2003; Grahne and Zhu, 2003; Shabtay et al., 2021; Shahbazi and Gryz, 2022). This data structure is designed to compactly represent transactions by compressing the database.

An FP-tree is constructed in the following steps:

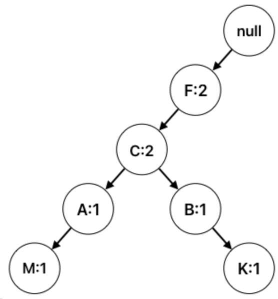
1. **Scan the Dataset:** The transaction database is scanned to determine the count of each item.
2. **Order Items:** Items in transactions are sorted in descending order of item counts.
3. **Build the Tree:** The FP-tree is built by reading each transaction and mapping it to a path in the tree, ensuring common prefixes are shared to compress the data.

Table 1: Initial Transactions.

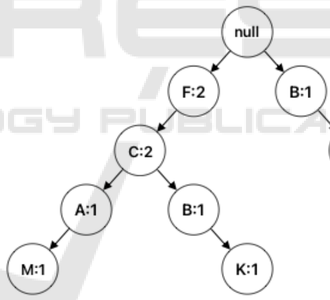
Transaction ID	Sorted items
1	F, C, A, M
2	F, C, B, K
3	B, E
4	F, C, A, M



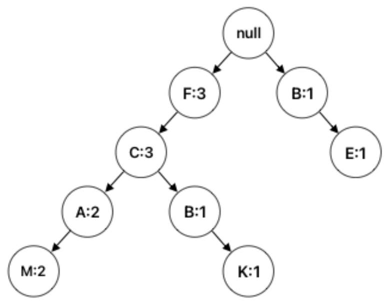
(a) Step 1



(b) Step 2



(c) Step 3



(d) Step 4

Figure 1: Progress of FP-tree construction from transactions in table 1.

Figure 1 demonstrates how the FP-tree structure is dynamically built using transaction from table 1, efficiently representing the frequent itemsets within the dataset.

FP-trees are particularly useful in applications where identifying frequent itemsets is crucial, such as market basket analysis, bioinformatics, and web usage mining. Their ability to efficiently handle large datasets makes them a powerful tool in data mining tasks. However, the potential of this data structure for storing association rules has not been fully explored.

4.2 Proposed Visualization Approach

To leverage the **FP-tree structure for visualizing association rules**, we propose a novel approach called the ”Trie of Rules.” This method adapts the FP-tree to effectively represent association rules, enabling users to comprehend the hierarchical relationships between items and the formation of rules while also reducing the size of the final plot by overlapping rules with common items.

Concept of Rules. In the Trie of Rules, each path from the root (Null node) to a node represents an association rule, where the nodes along the path form the antecedent, and the final node represents the consequent. Figure 2 illustrates the structure of a rule in the Trie of Rules. The item p is depicted as an element that exists in the trie but is not part of the evaluated rule $(f, c, a \rightarrow m)$. However, it can potentially become part of another rule. This structure allows users to trace hierarchical relationships between items, enhancing the interpretability and manageability of the visualization of the rules.

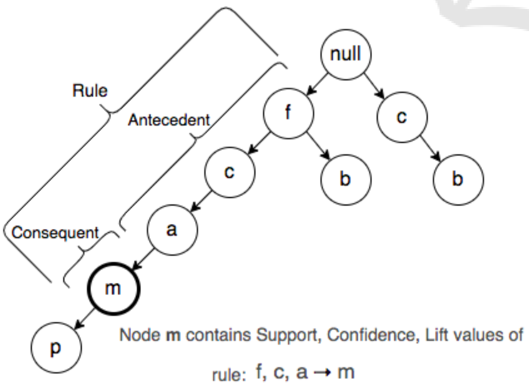


Figure 2: The structure of a rule in a Trie of Rules.

Metrics Display. Metrics are displayed through the color and size of nodes, and optionally, through the size of the caption near nodes. For instance, in Figure 4a, node size captures confidence while node color represents lift, although various other configurations are possible.

Our approach also facilitates the discovery of additional insights, such as clusters and substitute items:

- **Clusters:** Groups of items that frequently occur together can be easily identified through their shared paths in the FP-tree structure, revealing natural clusters within the data.
- **Substitute Items:** Items that can replace each other in transactions are revealed through the overlapping paths in the tree, providing insights into alternative itemsets.

4.3 Confidence for Compound Consequent

A unique feature of our approach is the ability to calculate confidence for rules with compound consequents directly from the graph structure. The confidence of a compound-consequent rule can be calculated as the multiplication of confidence values of the nodes in the consequent, as illustrated in Figure 3.

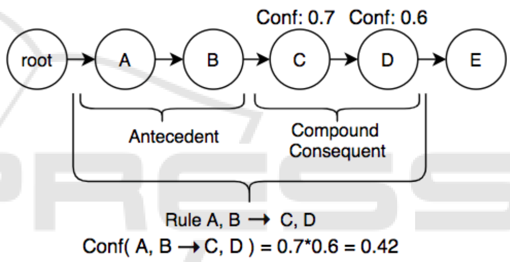


Figure 3: A rule with a compound consequent.

Although this method specifically applies to confidence, the support value for items with a compound consequent does not require additional calculation. The support of a rule $A, B, C \rightarrow D$ is equal to the support of the rule $A, B \rightarrow C, D$, as both rules refer to the same set of item occurrences within the dataset. Since the support measures the co-occurrence of items, the support for both rules remains the same. However, it is important to note that while the support is identical, the confidence differs. The confidence of $A, B \rightarrow C, D$ is based on how often C, D appear given A, B, whereas the confidence of $A, B, C \rightarrow D$ is calculated based on how often D appears given A, B, C.

The example rule in Figure 3 is part of a longer path within the trie, but we extract this portion to demonstrate that any section of the path can be taken as a rule. The figure also shows the item E, which exists in the trie but is not part of the current rule.

4.4 Case Study

For the implementation and testing of the Trie of Rules methodology, we used the ”Online Retail Logs”

(a) (b) Section A

Figure 4: (a) Trie of Rules visualization of the ARM results for the online retail dataset without captions displayed. (b) Zoomed section A of Figure 4a. LB stands for Lunch Bag.

dataset (Chen, 2015). This dataset, characterized by its large size and sparsity, contains 3,663 unique items and 18,484 transactions. The minimum support threshold for the ARM algorithm was set to 0.015, resulting in 234 association rules. We used the FP-growth algorithm (Han et al., 2000) to process the dataset and our developed library (implementation of the Trie of Rules methodology[1]) to produce the graph file.

The resulting Trie of Rules was visualized as a graph structure using Gephi 0.9.2 (Bastian et al., 2009). The default overlay method "Yifan Hu" (Hu, 2006) in Gephi was applied to enhance the clarity of the visualization.

Figure 4a illustrates the Trie of Rules generated from the Online Retail dataset. The visualization highlights clusters, the hierarchical structure of association rules, and substitute items, providing valuable insights into the dataset.

There are several valuable implications we can draw from exploring Figure 4b:

- The branch that starts with *LB RED* forms various rules that consist solely of Lunch Bag (LB) items of different designs: Vintage, Pink Polkadot, Cars Blue, etc. We can infer that these bags are often bought together in various designs. Based on this, we can propose selling these items as sets. Moreover, sets of color palettes can be formed based on the association rules observed in the Trie of Rules, for example, $(RED, VINTAGE)$ or $(RED, SUKI\ DESIGN, PINK\ POLKADOT)$. Given that *LB RED* starts this branch, we can imply that *LB RED* is the most popular and could be the "default" item in these sets.
- The branch that starts with *PINK TEACUP* creates several strong rules in the dataset. The color and size of the nodes indicate high Lift and Confidence values. However, this branch forms just two rules:

1. $PINK\ TEACUP \rightarrow GREEN\ TEACUP$
2. $(PINK\ TEACUP, GREEN\ TEACUP) \rightarrow ROSES\ TEACUP$

The first rule is a sub-rule of the second. We can imply that these items are often bought together with high probability. As with the previous branch, we can propose selling these items as sets of various designs. In this case, only one color palette can be proposed: $(PINK, GREEN, ROSES)$.

[1] https://github.com/ARM-interpretation/Trie-of-rules

5 EVALUATION

Evaluating visualization approaches for Association Rule Mining (ARM) is a complex task. Previous studies have employed various methods to assess the effectiveness of visualization techniques:

- Some researchers simply invite one or two experts to provide subjective feedback on their method's effectiveness (Menin et al., 2021; Varu et al., 2022).
- Others demonstrate the utility of their visualization techniques using "validation through awesome example" (Ong et al., 2002; Leung and Carmichael, 2009).
- Another common approach is to outline the advantages and disdavantages of the proposed methods without conducting rigorous user studies (Fernandez-Basso et al., 2019; Jentner et al., 2019; Hahsler and Chelluboina, 2011; Fister et al., 2023).

However, those methods are not considered as robust enough and objective; literature suggests using more comprehensive evaluation methodologies, such as those described by (Elmqvist and Yi, 2012), emphasising the importance of assessing cognitive load and user efficiency, especially when dealing with complex visualization tasks. Cognitive load refers to the amount of cognitive resources required to perform a task. As highlighted by (Yoghourdjian et al., 2021; Henike et al., 2020; Huang et al., 2009), it provides a quantitative measure to compare the efficiency of different visualization methods, making cognitive load a suitable metric in our study. A conceptual construct of cognitive load in the context of visualization efficiency (Huang et al., 2009) is illustrated in Figure 5.

Our evaluation focuses on measuring efficiency and learnability, similar to the approach used by (Huang et al., 2009). The evaluation process involved a carefully designed survey and tasks, structured as follows.

5.1 Survey Construction

We conducted a survey, which was approved by the ethical committee of [University Name]. The participants, 34 individuals with higher education backgrounds, completed the survey remotely on their own computers. We utilized the LimeSurvey platform to collect their responses and to record the time taken to answer each question. Participants were informed that their response times were being tracked.

Although the survey was anonymous, we ensured a diverse pool by using surveyswap.io, limiting potential participants to those with higher education in technical fields. Additionally, 14 participants were second-year computer science students from [University Name], consisting of 9 females and 5 males. This approach provided a balanced demographic, enhancing the robustness and interpretability of the results.

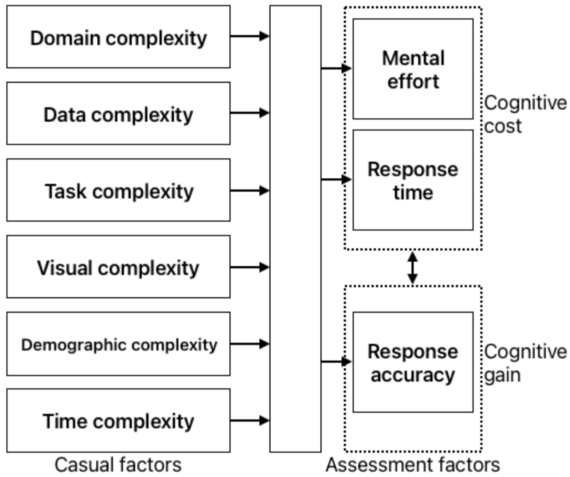


Figure 5: The construct of cognitive load for visualization understanding.

The survey took approximately 50 minutes for each participant and included four sections, one for each type of visualization: scatter plot, matrix-based, graph-based, and our proposed Trie of Rules approach. The sections were presented in a random order for each participant. At the beginning of the survey, participants were given a short introduction to ARM to ensure they could perform the given tasks.

Each section contained 9 questions:

- One introductory question to assess the ease of understanding the visualization method on a scale from 1 to 10, measuring learnability.
- Four simple questions focusing on tasks such as finding the support or confidence of a rule and identifying the rule with the maximum support or confidence.
- Four complex questions requiring deeper analysis, such as determining relationships between rules, identifying substitute items, assessing clusters, counting rules with a specific item, and finding the longest rule.

Participants were not limited in time and were asked the same questions across different visualization methods but with varying items to ensure consistency.

5.2 Measured Metrics

The following metrics were measured to evaluate the effectiveness of the visualization techniques:

- **Response Time (RT):** The time taken to complete each task. Shorter response times indicate more efficient visualizations.
- **Response Accuracy (RA):** The correctness of the answers provided. Higher accuracy indicates more effective visualizations.
- **Mental Effort (ME):** Self-reported effort on a scale of 1 to 10. Lower mental effort suggests that the visualization is easier to understand and use.

To standardize the results and facilitate a fair comparison across different visualization methods, we calculated z-scores for these metrics following the methodology proposed by (Huang et al., 2009). The z-score transformation normalizes the data by subtracting the mean and dividing by the standard deviation of the respective metric, resulting in a standardized score with a mean of 0 and a standard deviation of 1. The formula for calculating the z-score is:

$$z = \frac{X - \mu}{\sigma}$$

where X is the raw score, μ is the mean of the scores, and σ is the standard deviation.

We used the following formula for visualization efficiency:

$$E = Z_{RA} - Z_{ME} - Z_{RT}$$

In this formula, E represents the efficiency via cognitive load, Z_{RA} is the z-score for response accuracy, Z_{ME} is the z-score for mental effort, and Z_{RT} is the z-score for response time. This metric captures the trade-off between accuracy, effort, and time, providing a comprehensive measure of visualization efficiency. High efficiency is achieved when high accuracy is associated with low mental effort and short response time.

5.3 Survey Results and Analysis

The results of our evaluation are summarized in Table 2 and Table 3.

In terms of accuracy, the Trie of Rules method demonstrated better performance on complex questions (0.59) compared to the other methods (Matrix: 0.17, Graph: 0.29, Scatter: 0.23). This indicates that while the Trie of Rules may be novel and less familiar to users, its structured representation of association rules enables more accurate analysis of complex relationships. However, for simple questions, the accuracy of the Trie of Rules (0.44) was on par with the Scatter plot (0.44) and better than the Matrix (0.34) and Graph (0.20) methods. This suggests that while the Trie of Rules is effective for both simple and complex tasks, its advantage becomes more pronounced with increased complexity.

Regarding mental effort, all methods showed no significant difference, as indicated by the ANOVA test results (p-value < 0.05). This indicates that the complexity of the questions impacted time and accuracy rather than mental effort. The Scatter plot required the least effort (2.57), probably because it is the most familiar and commonly used scientific visualization method. The Trie of Rules method showed moderate mental effort (3.11), indicating that while it is a novel approach, it is not significantly more challenging to understand and use compared to existing methods.

The response time for simple questions was slightly higher for the Trie of Rules (56 seconds) compared to the other methods, with the Scatter plot being the fastest (40 seconds). This suggests that users may need more time to familiarize themselves with the Trie of Rules. However, for complex questions, the Trie of Rules (35 seconds) performed on par with the Scatter plot (35 seconds), indicating that once users become familiar with the method, they can analyze complex information just as quickly as with more traditional methods.

5.4 Discussion

The results indicate that the Trie of Rules method offers a significant advantage in terms of accuracy and efficiency, particularly for complex questions, while maintaining a moderate mental effort comparable to existing methods.

The slightly higher response time for simple questions indicates that there is a learning curve associated with the Trie of Rules. This could be due to its novel representation compared to more familiar visualization methods like the Scatter plot. However, the improved accuracy and efficiency for complex questions highlight the potential benefits of this method, especially in scenarios where users need to analyze intricate relationships within the data.

Furthermore, the findings suggest that the benefits of the Trie of Rules may become more apparent with larger datasets and more complex association rules. Future studies could explore the impact of different dataset sizes and structures on the effectiveness of the Trie of Rules. For instance, with twice the number of data points, the advantages of the Trie of Rules in handling complex information efficiently might be even

Table 2: Means of response time, accuracy, mental effort, and efficiency on simple questions.

	Trie of Rules	Matrix	Graph	Scatter
Time (sec.)	56.00	43.00	73.00	40.00
Accuracy	0.44	0.34	0.20	0.44
Effort	3.11	3.32	3.03	2.57
Efficiency	0.23	-0.46	-2.76	2.99

Table 3: Means of response time, accuracy, mental effort, and efficiency on complex questions.

	Trie of Rules	Matrix	Graph	Scatter
Time (sec.)	35.00	40.00	46.00	35.00
Accuracy	0.59	0.17	0.29	0.23
Effort	3.11	3.32	3.03	2.57
Efficiency	1.89	-1.99	-1.56	1.66

more pronounced.

Overall, the Trie of Rules method demonstrates promising potential for enhancing the interpretability and usability of ARM visualizations. By offering a structured and efficient way to represent association rules, it can help users uncover hidden patterns and relationships within large datasets, ultimately facilitating better decision-making and knowledge discovery. Future work will focus on developing software tools to facilitate the adoption of this methodology and further optimizing the user interface and experience to improve the efficiency of the visualization process.

6 CONCLUSION

Association Rule Mining is a valuable technique for uncovering hidden patterns in large datasets, and the efficiency of individuals interpreting these results is greatly influenced by the effectiveness of the visualization techniques employed. Existing visualization methods often struggle with scalability, interpretability, and the effective representation of rule structures, limiting their practical utility in real-world applications.

In this paper, we introduced a novel visualization technique called the ”Trie of Rules.” This method leverages the FP-tree structure to compactly and effectively represent association rules, addressing the common issues faced by traditional visualization approaches. Our approach not only captures a wealth of information and reveals implicit insights, such as clusters and substitute items, but also maintains manageable visualization size by overlapping common items.

We conducted a comprehensive evaluation to compare the Trie of Rules with existing visualization methods through a survey measuring cognitive load. The results demonstrated that our method outperforms others in terms of efficiency, particularly in handling complex queries, while maintaining comparable learnability.

Our findings indicate that the Trie of Rules method significantly enhances the interpretability and usability of ARM visualizations. Future work will focus on developing software tools to facilitate the adoption of this methodology and further researching how user interface and user experience can be optimized to improve the efficiency of the visualization process.

REFERENCES

Agrawal, R., Imieliński, T., and Swami, A. (1993). Mining Association Rules Between Sets of Items in Large Databases. In *ACM SIGMOD Record*, volume 22, pages 207–216.

Alyobi, M. A. and Jamjoom, A. A. (2020). A visualization framework for post-processing of association rule mining. *International Journal Transaction on Machine Learning and Data Mining*, 2020(2):83–99.

Bastian, M., Heymann, S., and Jacomy, M. (2009). Gephi: An open source software for exploring and manipulating networks.

Bodon, F. and Rónyai, L. (2003). Trie: an alternative data structure for data mining algorithms. In *Mathematical and computer modelling*, volume 38, pages 739–751.

Buono, P. and Costabile, M. F. (2005). *Visualizing Association Rules in a Framework for Visual Data Mining*, pages 221–231. Springer Berlin Heidelberg, Berlin, Heidelberg.

Chen, D. (2015). Online Retail. UCI Machine Learning Repository. DOI: https://doi.org/10.24432/C5BW33.

Elmqvist, N. and Yi, J. S. (2012). Patterns for visualization evaluation. *ACM International Conference Proceeding Series*, (October 2012).

Ertek, G. and Demiriz, A. (2006). A framework for visualizing association mining results. In *Computer and*

Information Sciences – ISCIS 2006, pages 593–602. Springer Berlin Heidelberg.

Fernandez-Basso, C., Ruiz, M. D., Delgado, M., and Martin-Bautista, M. J. (2019). A comparative analysis of tools for visualizing association rules: A proposal for visualising fuzzy association rules. In *Proceedings of the 11th Conference of the European Society for Fuzzy Logic and Technology (EUSFLAT 2019)*, pages 520–527. Atlantis Press.

Fister, I., Fister, I., Fister, D., Podgorelec, V., Fister, I., and Salcedo-Sanz, S. (2023). A comprehensive review of visualization methods for association rule mining: Taxonomy, challenges, open problems and future ideas. *Expert Systems with Applications*, 233(June):120901.

Geng, L. and Hamilton, H. J. (2006). Interestingness measures for data mining: A survey. *ACM Comput. Surv.*, 38(3):9–es.

Grahne, G. and Zhu, J. (2003). Efficiently using prefix-trees in mining frequent itemsets. *Proc. of the 1st IEEE ICDM Workshop on Frequent Itemset Mining Implementations*, pages 236–245.

Hahsler, M. (2023). ARULESPY: Exploring Association Rules and Frequent Itemsets in Python. (Raschka 2018).

Hahsler, M. (2024). A Probabilistic Comparison of Commonly Used Interest Measures for Association Rules.

Hahsler, M. and Chelluboina, S. (2011). Visualizing Association Rules: Introduction to the R-extension Package arulesViz. Technical Report February.

Hahsler, M., Chelluboina, S., and Hornik, D. (2017). Visualizing association rules: Introduction to the r-extension package arulesviz. *Journal of Statistical Software*, 83(1).

Han, J., Pei, J., and Yin, Y. (2000). Mining frequent patterns without candidate generation. *SIGMOD Record (ACM Special Interest Group on Management of Data)*, 29(2):1–12.

Han, J., Pei, J., Yin, Y., and Mao, R. (2004). Mining frequent patterns without candidate generation: A frequent-pattern tree approach. *Data Mining and Knowledge Discovery*, 8(1):53–87.

Henike, T., Kamprath, M., and Hölzle, K. (2020). Effecting, but effective? How business model visualisations unfold cognitive impacts. *Long Range Planning*, 53(4).

Hofmann, T. and Buhmann, J. M. (2000). Multidimensional scaling and data clustering. In *Advances in Neural Information Processing Systems*, pages 459–466.

Hu, Y. (2006). The Mathematica ® Journal Efficient, High-Quality Force-Directed Graph Drawing. *Methematica Journal*, 10:37–71.

Huang, W., Eades, P., and Hong, S. H. (2009). Measuring effectiveness of graph visualizations: A cognitive load perspective. *Information Visualization*, 8(3):139–152.

Jentner, J., Heitmann, B., and Nagel, W. E. (2019). A survey on visualization for mining association rules and frequent item sets. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 9(2).

Jr., R. J. B., Agrawal, R., and Gunopulos, D. (1999). Constraint-based rule mining in large, dense databases. In *Proceedings of the 15th International Conference on Data Engineering*, pages 188–197.

Klemettinen, M., Mannila, H., Ronkainen, P., Toivonen, H., and Verkamo, A. I. (1994). Finding interesting rules from large sets of discovered association rules. In *Proceedings of the Third International Conference on Information and Knowledge Management*, pages 401–407.

Leung, C. K. S. and Carmichael, C. L. (2009). FpViz: A visualizer for frequent pattern mining. *Proceedings of the ACM SIGKDD Workshop on Visual Analytics and Knowledge Discovery, VAKD '09*, (January 2009):30–39.

Luna, J. M., Ondra, M., Fardoun, H. M., and Ventura, S. (2018). Optimization of quality measures in association rule mining: an empirical study. *International Journal of Computational Intelligence Systems*, 12:59–78.

Menin, A., Cadorel, L., Tettamanzi, A., Giboin, A., Gandon, F., and Winckler, M. (2021). ARViz: Interactive Visualization of Association Rules for RDF Data Exploration. In *Proceedings of the International Conference on Information Visualisation*, volume 2021-July, pages 13–20. IEEE.

Ong, K.-h., Ong, K.-l., Ng, W.-k., and Lim, E.-p. (2002). CrystalClear: Active Visualization of Association Rules. *ICDM'02 International Workshop on Active Mining AM2002*, (February):1–6.

Rainsford, C. P. and Roddick, J. F. (2000). Visualisation of temporal interval association rules. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, volume 1983, pages 91–96.

Shabtay, L., Fournier-Viger, P., Yaari, R., and Dattner, I. (2021). A guided fp-growth algorithm for mining multitude-targeted item-sets and class association rules in imbalanced data. *Information Sciences*, 553:353–375.

Shahbazi, N. and Gryz, J. (2022). Upper bounds for can-tree and FP-tree. *Journal of Intelligent Information Systems*, 58(1):197–222.

Shaukat Dar, K., Zaheer, S., and Nawaz, I. (2015). Association rule mining: An application perspective. *International Journal of Computer Science and Innovation*, 1:29–38.

Varu, R., Christino, L., and Paulovich, F. V. (2022). ARMatrix: An Interactive Item-to-Rule Matrix for Association Rules Visual Analytics. *Electronics (Switzerland)*, 11(9).

Wu, T., Chen, Y., and Han, J. (2010). Re-examination of interestingness measures in pattern mining: A unified framework. *Data Mining and Knowledge Discovery*, 21(3):371–397.

Yoghourdjian, V., Yang, Y., Dwyer, T., Lawrence, L., Wybrow, M., and Marriott, K. (2021). Scalability of Network Visualisation from a Cognitive Load Perspective. *IEEE Transactions on Visualization and Computer Graphics*, 27(2):1677–1687.