Article

ARMatrix: An Interactive Item-to-Rule Matrix for Association Rules Visual Analytics

Rakshit Varu [ID], Leonardo Christino [ID] and Fernando V. Paulovich *[ID]

Faculty of Computer Science, Dalhousie University, 6050 University Avenue, Halifax, NS B3H 4R2, Canada; rk996644@dal.ca (R.V.); christinoleo@dal.ca (L.C.)
* Correspondence: paulovich@dal.ca; Tel.: +1-902-494-1986

Abstract: Amongst the data mining techniques for exploratory analysis, association rule mining is a popular strategy given its ability to find causal rules between items to express regularities in a database. With large datasets, many rules can be generated, and visualization has shown to be instrumental in such scenarios. Despite the relative success, existing visual representations are limited and suffer from analytical capability and low interactive support issues. This paper presents ARMatrix, a visual analytics framework for the analysis of association rules based on an interactive item-to-rule matrix metaphor which aims to help users to navigate sets of rules and get insights about co-occurrence patterns. The usability of the proposed framework is illustrated using two user scenarios and then confirmed from the feedback received through a user test with 20 participants.

Keywords: visual analytics; association rules; matrix-based visualization

1. Introduction

Data mining or machine learning are vast fields, with growing importance in the last few decades. In general, data mining or machine learning techniques can be split into two major groups, supervised and unsupervised approaches or a combination of both. The difference is that the former learns from existing data to create a model, and the latter creates models to discover patterns without associated information (e.g., labels) [1]. Once unsupervised tasks have no ground truth, it is essential to support users to interpret and make sense of the produced results. Otherwise, the abundance of possible distinctive (and still correct) results make the usage of such techniques impractical [2].

One popular unsupervised approach is Association Rule Mining (ARM) [3]. ARM aims to find frequent relationships between variables in a dataset, deriving causal rules to express regularities [3] with application varying from market data [4], intrusion detection [5], bioinformatics [6], to name a few. Despite all the benefits of ARM, some issues need to be addressed to make it useful, principally the usual discovery of significant rule counts, since, as the rule number increases, it becomes more difficult to inspect the results and find relevant relationships using simple textual representations [7].

One possible solution is visualization [8], using the human visual system to improve our capacity to analyze large quantities of information. In this direction, different representations have been proposed to support the analysis of ARM results. Table-based techniques arrange textual descriptions of rules in a table format [9–11] (spreadsheet), but are only applicable when the type of analysis is simple to execute, although, in practice, this is the standard metaphor for commercial systems [12]. Graph-based techniques [9,10,13–15] is an advance compared to it, but analytical capability is an issue since the "if-then" rules structure is lost in the visual representations. Matrix-based showed to be a good solution for these problems. However, the existing proposed matrix organizations make the visualization too dense to "read" [16] or lack the fine support for rule comparison [17]. 3D matrix visualizations [18–21] have been suggested to address this, but without proper interaction

resources, users might have to use their imaginations to understand the results or might easily misinterpret them [22].

This paper proposes *ARMatrix*, a visual analytics framework based on a 2D matrix metaphor to address these issues, especially the detailed comparison between rules without losing the notion of "if-then" relationships. *ARMatrix* arranges rules as columns and items (attributes) as rows and provides different mechanisms to order and filter the resulting layout to allow the execution of different analytical tasks and focus on sub-sets of items of interest.

More specifically, the contributions of *ARMatrix* are:

- a framework to support the analysis of association rules and the interpretation of the produced results;
- a novel item-to-rule matrix visualization to support a detailed comparison between rules without losing the notion of "if-then" relationships;
- user test evaluation showing that, compared to the usual table-based representation employed by commercial systems, participants were able to produce more precise results using our solution while feeling more confident about the attained conclusions.

The remainder of this paper is structured as follows. Section 3 presents background information about ARM and the related interest measures. Section 2 covers the literature in ARM visualization and discusses its gaps. Section 4 presents the design goals we followed along with a detailed description of our solution. Section 5 presents two use-cases to show how to use *ARMatrix* in practice and a user study comparing it against the usual strategy employed by commercial systems. Finally, Section 6 discusses our limitations, and Section 7 delineates our conclusions and future work.

2. Related Work

When it comes to visual analytics or visual data mining processes, visualization can assist computational models in three different stages [23]: before the data are processed to create a model, supervise and steer the run of algorithms, and interpret the produced results. The existing solutions for the visualization of association rules focus on the last two and can be divided into table-based, graph-based, matrix-based, and 3D visualization techniques [24].

Table-Based Techniques. Table-based techniques are one of the simplest methods to display association rules. In this type of visualization, table headers consist of rule names and interesting measures (support, confidence, and others), with cells containing textual rules and computed values. When integrated with other visualizations like dimensionality reduction layouts, tables are typically used to explore rules as a secondary visualization [9–11].

The table form is easy to utilize when the dataset and the rules are small in number but gets time-consuming and challenging to interpret when the volume is high. Comparing rules or analyzing groups of rules is laborious, and ordering rules based on different measures or displaying more information in a confined space is difficult. However, in practice, the table-based metaphor is still one of the most common approaches to analyze the result of association rules and can be found in several commercial systems [12].

Graph-Based Techniques. Graph visual representations have also been used to display association rules. Typically, items are represented as nodes, and the edges denote rules' properties [13]. Yamamoto and Oliveira [14] use a force-directed algorithm based on the spring metaphor to draw item graphs, where the force between nodes (edges) is proportional to the frequency of the items occurring together. The visualizations generated in [9,10] are also represented using item graphs but with connections depicting interesting measures. The work of Klemettinen et al. [15] employs a similar idea along with bar charts to browse rules. CrystalClear [25] simplifies the process and uses a tree visualization to show a hierarchical view. Castillo-Rojas et al. [12] also uses a hierarchical structure as primary visualization and other visual representations as detailed visual elements.

Recently, other line or graph-based techniques have been suggested. In [26], differential evolution is taken into consideration for the generation of association rules over time. These rules are then displayed using Sankey diagrams where entities are nodes, and the edges of different widths are the weights calculated from the attributes (support and confidence). The temporal nature of the rules is tried to be addressed in their technique, but the problem stands as the number of rules increases and the placement of nodes within layers becomes unmanageable. ARViz [27] for RDF data exploration offers three different visual representations, namely a scatter plot for giving the overview of the rules, a chord diagram to display subsets of rules, and an association graph for the exploration of items. They offer a detail-on-demand method to show additional information and an interactive prompt for exploring and discovering new rules. Though the circular paginated visualization is helpful to explore subsets, it has a limit of 150 rules to be displayed without getting the rules jumbled. Alyobi and Jamjoom [28] tried to address the scalability issue for visualizing rules using the collapsed tree visualization, where they clustered the rules by groups of antecedents together based on the lift value through the K-means clustering technique. The primary layer of the tree represents the antecedents, while the expanded subtree indicates the consequent and their association with the predecessor. They used color to differentiate clusters with different levels of lift. On-demand detail is provided when the user interacts with the tree. The technique can distinguish clusters easily but fails to address the issue of comparing the rules to one another.

As in table-based techniques, the main drawback of graph-based methods is visual scalability. When the rule number is large, some rules can coincide with one another, resulting in information occlusion. In addition, node-link-based graph visualizations can result in inter-twinned connections, which gets challenging to interpret, and the "if-then" rules structure is lost since the focus is on items. Our approach helps to visualize all the rules without making them difficult to understand, organizing all the information using a non-overlapping matrix metaphor where the relationships between antecedents and consequents, therefore, the rules are explicit.

Matrix-Based Techniques. Matrix is also a usual metaphor to display association rules and convey different information depending on its organization. In the item-to-item organization [16], rows and columns consist of items representing the antecedents and consequents, while cells are colored using a measured value for the rule. In the itemset-to-itemset organization [17], rows represent antecedent itemsets and the columns, consequent itemsets (or vice-versa), while cells are colored to show different rules based on lift.

The item-to-item matrix visualization does not incorporate a way to deal with repeated rules. Usually, the visualization becomes too dense, and it is challenging to "read" the rules. The itemset-to-itemset organization is a compact version of the item-to-item, where more rules can be depicted. However, comparison of rules or groups of rules, especially if rules share items in the antecedent and consequent, is challenging. Notice that similar matrices have also been used to represent functional dependencies between attributes [29], which could be adapted to an item-to-item representation if the attributes are considered antecedents and consequents of association rules. However, such adaptation would still present these same limitations. Our approach employs a different organization, item-to-rule, explicitly mapping the rules per item, easing the process of comparing rules, and analyzing groups of rules with shared items.

3D Techniques. Three-dimensional visualizations have also been suggested for the analysis of association rules. In general, the core concept is to use any visual metaphor mentioned before and represent rules with extra information using the third dimension [19–21]. Couturier et al. [18] use a matrix to represent clusters of rules and then provide support for analysis using 3D bar charts. Wong et al. [30], and Chakravarthy and Zhang [31] also use a 3D matrix-based visualization to show rules in the item-to-rule base. These proposed visualizations do not deliver the user with the chance to change the visualizations interactively. In addition, the 3D visualization requires intensive manipulation for results' interpretation, which can get overwhelming. If not provided with proper interaction re-

sources, the user might have to use their imaginations to understand the results or might easily misinterpret them [22]. Our approach also follows an item-to-rule organization, but it is a 2D representation that offers different interactive mechanisms to allow a detailed exploration of rules.

Alternative visual metaphors that do not fit in these categories have also been proposed, for instance, using scatterplots of support vs. confidence [25,32] or dimensionality reduction layouts [33] to explore items or rules relationships, lacking the ability to understand "if-then" associations. Treemap-like visualizations have also been proposed to analyze rules [34], but it is only suitable for a very small number of rules. Our focus is on larger sets of rules with aid to the direct interpretation of "if-then" associations between rules and groups of rules.

3. Background

Association rules are "if-then" statements that show how often things occur together within a database [3]. Let $\mathcal{I} = i_1, i_2, \ldots, i_N$ be a set of items and $\mathcal{T} = t_1, t_2, \ldots, t_M$ a set of transactions, where each transaction t consists of a set of items $t \subseteq \mathcal{I}$. Association rule mining algorithms help find relations between itemsets, that is, collections of items, in the form of $\mathcal{A} \Rightarrow \mathcal{C}$, where $\mathcal{A}, \mathcal{C} \subseteq \mathcal{I}$ and $\mathcal{A} \cap \mathcal{C} = \varnothing$. $\mathcal{A}$ is called antecedent (if) and $\mathcal{C}$ consequent (then), indicating that whenever $\mathcal{A}$ occurs, $\mathcal{C}$ also occurs with a certain probability. For instance, for a supermarket transaction database containing what customers bought together, the rule $\{\text{Diaper}, \text{Milk}\} \Rightarrow \{\text{Beer}\}$ implies that when a customer purchases diapers and milk, s/he also tends to purchase beer [1].

Depending on factors such as performance and memory consumption, association rules can be extracted using different algorithms, including Apriori, FP-growth, and Eclat [35]. In common, to control the number and relevance of the produced rules, two measures are used, support and confidence [3]. Based on the support of an itemset $\mathcal{B} \subseteq \mathcal{I}$, which is the ratio of the number of transactions that contains $\mathcal{B}$ by the total number of transactions in the database, that is

$$supp(\mathcal{B}) = \frac{|t \in \mathcal{T}; \mathcal{B} \subseteq t|}{|\mathcal{T}|}. \tag{1}$$

The support of a rule $\mathcal{A} \Rightarrow \mathcal{C}$ is defined as the ratio of transactions having $\mathcal{A}$ as antecedent and $\mathcal{C}$ as consequent by the total number of transactions in the database

$$supp(\mathcal{A} \Rightarrow \mathcal{C}) = supp(\mathcal{A} \cup \mathcal{C}). \tag{2}$$

Support ranges in $[0, 1]$, the larger the better. If a rule has the support of 0.1, it means that 10% of the total rules contain $\mathcal{A}$ and $\mathcal{C}$.

Confidence is derived from support and is the ratio of the number of transactions having both $\mathcal{A}$ and $\mathcal{C}$ to the number of transactions that contain $\mathcal{A}$. Confidence indicates how many times if $\mathcal{A}$ appeared as antecedent, $\mathcal{C}$ appeared as consequent, and is given by

$$conf(\mathcal{A} \Rightarrow \mathcal{C}) = \frac{supp(\mathcal{A} \Rightarrow \mathcal{C})}{supp(\mathcal{A})}. \tag{3}$$

Confidence ranges in $[0, 1]$, the larger the better. The strength of the rule depends upon this measure. If a rule has the confidence of 0.9, it means that 90% of the transactions containing $\mathcal{A}$ also contain $\mathcal{C}$.

Besides support and confidence, the extracted rules' relevance can be computed using other interesting measures [36]. As explained, confidence measures the probability of antecedent and consequent occurring together. However, it does not consider the correlation between items. To overcome this issue, the lift measure can be used [37]. Lift

helps to find how often $\mathcal{A}$ and $\mathcal{C}$ appear together if they are statistically independent and is given by

$$lift(\mathcal{A} \Rightarrow \mathcal{C}) = \frac{conf(\mathcal{A} \Rightarrow \mathcal{C})}{supp(\mathcal{C})}. \tag{4}$$

Lift ranges in $[0, \infty]$. If $\mathcal{A}$ and $\mathcal{C}$ are independent, lift is 1. If it is greater than 1, it indicates the dependence of the items and can be used to predict consequent in unseen cases. For visualization purposes, we prefer to use the standardized lift [38] to avoid problems when mapping infinity values, given by

$$lift'(\mathcal{A} \Rightarrow \mathcal{C}) = \frac{lift(\mathcal{A} \Rightarrow \mathcal{C}) - \lambda}{v - \lambda}, \tag{5}$$

where

$$\lambda = \frac{\max\{supp(\mathcal{A}) + supp(\mathcal{C}) - 1, 1/|\mathcal{T}|\}}{supp(\mathcal{A})supp(\mathcal{C})}$$
$$v = \frac{1}{\max\{supp(\mathcal{A}), supp(\mathcal{C})\}}. \tag{6}$$

The standardized lift ranges in $[0, 1]$. Values close to zero indicate random co-occurrences of items while 1 indicates the highest dependence between them.

4. ARMatrix

This section presents *ARMatrix*, an item-to-rule matrix-based visualization framework to assist with association rule analysis. *ARMatrix* was designed to support the most critical analytical tasks found in the literature and address the existing solutions' major limitations (Section 2). The design goals we follow are listed below.

G1—Visual scalability. Usually, the number of extracted association rules is high, and as the number increases, the analysis becomes more difficult [7,24]. Our primary design goal is to provide a representation that can visually scale without making it difficult to understand relationships between rules' antecedents and consequents by better utilizing the available visual space, only being limited by the user's display size and resolution.

G2—Comparative analysis. Comparing rules or groups of rules that share some property is the most common task of association rules analysis [15,19,32]. Our second goal is to support comparative analysis to reveal differences between rules regarding antecedents, consequents, and interesting measures.

G3—Interactive detailed exploration. Given the usual number of mined rules, our third goal is to provide interactive mechanisms to help users extract relevant insights and focus the analysis on the information of interest and get details on demand [16].

4.1. Overview

To support the above design goals, *ARMatrix* is implemented as an interactive framework composed of three different panels, presented in Figure 1. The *Control* panel allows users to set the parameters for the rule extraction and arrange the visualization according to interesting measures and retrieve analysis states Ⓐ. The *Visualization* panel is composed of two different visual representations to support rules' investigation. A novel item-to-rule matrix representation to show the extracted rules, as composed by antecedent and consequent, and a radar chart to allows for comparison and detailed inspection Ⓑ. Finally, the *Filtering* panel helps to find subsets of rules of interest based on items to focus the analysis Ⓒ.

The usual workflow for using *ARMatrix* starts with rules generated considering thresholds of support and confidence. The resulting rules are then displayed in the item-to-rule matrix, where users can analyze rules to verify the frequency of items, antecedent-

consequent relations, and the difference between rules based on the interesting measures. Hovering on an item highlights the rule associated with it and shows measures related to the rule. During the analysis, users can re-order the rules and items based on different criteria and find subsets of interest by selecting different filter options. By double-clicking on different rules, users can compare rules' different interesting measures using a radar chart. The whole process can be iterated multiple times depending on the analysis requirements. During this process, the sets of filtered rules can be stored for later use. The saved states can be deleted or retrieved at any time.
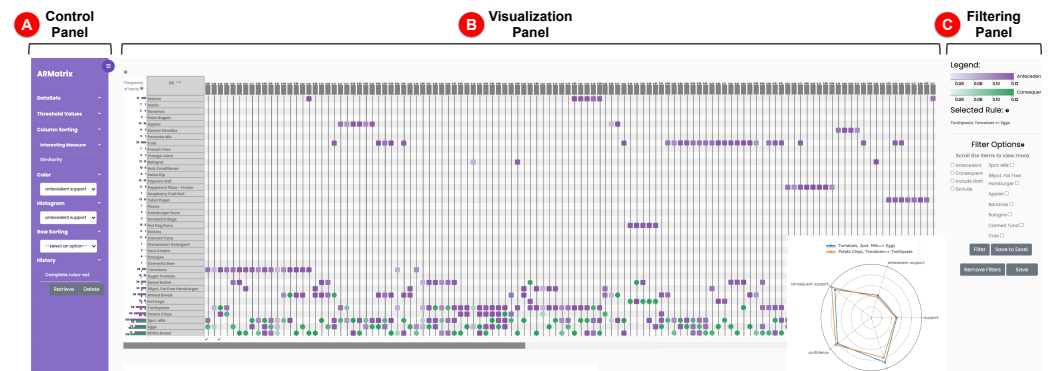

Figure 1. *ARMatrix* framework interface and its panels. The *Visualization* panel (**B**) is composed of two visual representations, a item-to-rule matrix to support the analysis of rules antecedent-consequent relations and a radar layout for detailed investigation and rules comparison. The *Control* panel (**A**) allows users to set the parameters for the rule extraction and arrange the visualization according to interesting measures, and the *Filtering* panel (**C**) helps users to select rules/items of interest and focus the analysis.

ARMatrix is implemented as an interactive web application using python in the backend [39] and D3 [40], Plotly Javascript and JQuery in the front-end. A representation of this architecture can be seen in Figure 2. Association rules are extracted and used for visualization through the *ARMatrix* dashboard, and the available interactions (in green) trigger database operations (in yellow) to update the visualization. For the rule extraction, we opt to use the FP-growth over the Apriori algorithm to reduce waiting times, but our system is algorithm agnostic, and any algorithm could be used.
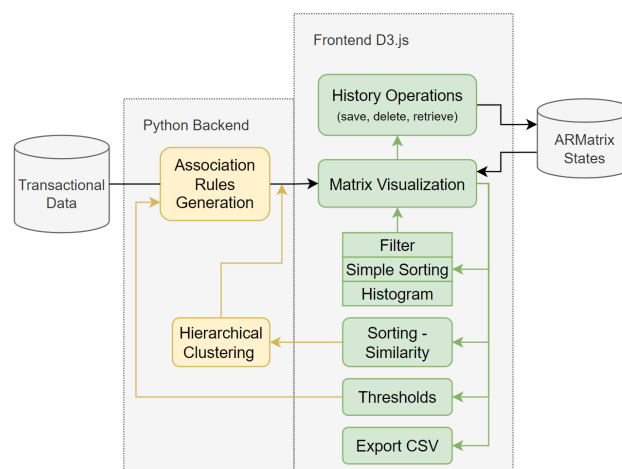

Figure 2. *ARMatrix* framework architecture. The system uses transactional data for the creation of a AR rules during the visualization task. These rules are visualized through the *ARMatrix* dashboard, and the available interactions (in green) trigger database operations (in yellow) to update the visualization.

4.2. Visualization Panel

The main *ARMatrix* visualization panel includes two components. An item-to-rule matrix representation to display rules and a radar chart to compare rules considering different interesting measures. In the matrix visualization, each row represents an item $i \in \mathcal{I}$, and each column represents a rule $\mathcal{A} \Rightarrow \mathcal{C}$ with $\mathcal{A}, \mathcal{C} \subseteq \mathcal{I}$ consisting of both antecedent $\mathcal{A}$ and consequent $\mathcal{C}$. The items in a single rule are connected with a black line at the center of the column to guide users when analyzing rules, and matrix rows' backgrounds are alternatively filled in gray and white to guide users when analyzing items. Color and shape are used to identify antecedent and consequents. Antecedents are represented by purple rounded squares and consequents by green circles. Tooltips containing rules' interesting measures are shown upon rolling the mouse over.

In the resulting matrix, selected interesting measures can be mapped to color brightness, the darker the color, the larger the value, and as histograms on top of each matrix column (rule) so that pair-wise comparison can be executed between measures. In addition, the frequencies of each item as antecedent and consequent are mapped to histograms on the matrix's left side. Purple color bars represent antecedent frequency, and green color bars represent consequent frequency. A detailed example of the resulting visualization is presented in Figure 3. This item-to-rule matrix compact representation allows for displaying a large number of rules (goal **G1**), supporting the comparison of rules based on antecedents and consequents (goal **G2**). If the number of rules or items is larger than the horizontal or vertical space, the matrix can be zooming out to allow for a general overview, and horizontal and vertical scroll bars are added. However, in general, association rules analysis is based on subgroups of rules, focusing on items or sets of items of interest.
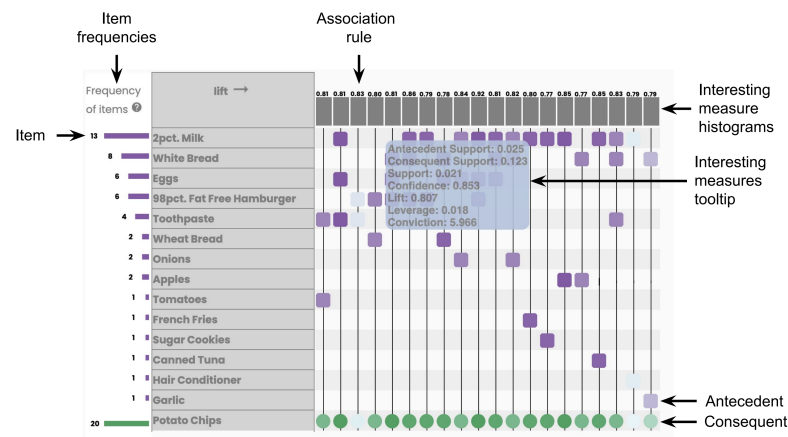

Figure 3. The proposed new item-to-rule matrix visual representation. Rows represent items and columns rules. Antecedents are represented by purple round squares and consequents by green circles. Interesting measures can be mapped to color brightness, histograms on top of each rule, and to sort the columns (rules). Histograms in each row represent the frequency of items as consequent and antecedent, and rows can be ordered based on it.

If users want to compare sets of rules considering all interesting measures and have a detailed analysis of individual rules, rules can be select by double-clicking the items in the column representing a rule. Once selected, a tick appears on the bottom of the selected column for indication. The chosen rules are then compared to one another using a radar chart where each anchor represents an interesting measure. The axes of the radar chart are normalized according to the selected rules to help in the comparison, but the original values are displayed by rolling the mouse over the radar layout anchors. Rules in textual format are also displayed for the selected rules. These features help examine and compare multiple rules together based on their measure values (goal **G3**). An example of the resulting visualization is presented in Figure 1 (bottom-right).

4.3. Rule Ordering and Filtering

The item-to-rule matrix can be transformed by sorting its rows and columns in different ways to ease the analysis. Initially, the matrix columns are sorted based on the decreasing value of their antecedent support from left to right. However, users can select other interesting measures to sort the columns in decreasing order (in this paper, we allow for support, confidence, and lift orderings). Rules can also be sorted based on similarity, allowing the analysis of groups of similar rules. In this process, we first create a binary matrix $\mathbf{A}_{P \times Q}$, where the pth row represents the rule $r_p = \mathcal{A}_p \Rightarrow \mathcal{C}_p$, the qth column represents the item $i_q \in \mathcal{I}$, and the entries a_{pq} indicates if i_q is part of rule r_p, that is

$$a_{pq} = \begin{cases} 1 & \text{if} i_q \in \{\mathcal{A}_p \cup \mathcal{C}_p\} \\ 0 & \text{otherwise.} \end{cases} \tag{7}$$

Once this matrix is created, the rules are sorted by applying a hierarchical clustering [1] algorithm in **A**, using the post-order traversal order of the resulting dendrogram (only the leaves are taken into consideration) to sort the columns. In this way, the most similar rules (or groups of rules) are placed together, and the order of similarity is decreased from left to right in the resulting matrix. To calculate the similarity between rules in **A**, we use the *Jaccard* binary distance so that rules composed by similar sets of items are deemed more related.

Rows can also be sorted in ascending/descending order based on the frequency of an item being an antecedent or consequent, so the most frequent items can be easily spotted. This flexible ordering mechanism allows for different arrangements of the item-to-rule matrix, enabling users to quickly discover, for instance, the more frequent antecedent or consequent items among the rules with high lift, shedding some light on what happens to one item (or set of items) if another item (or set of items) is removed from $\mathcal{I}$.

Besides ordering, the item-to-rule matrix can also be transformed by filtering rules to focus the analysis. This is done by selecting the items of interest, displaying only rules containing them as antecedent, consequent, or both. It is also possible to do the opposite, setting the irrelevant items, displaying only the rules that do not contain them. These filters help visualize the necessary subsets of the rules for analysis without losing the authenticity of the rule measures (goal **G3**). All the items are considered while calculating rules so that no item is excluded for rule calculation, and filters are applied only for visualization purposes.

All the ordering and filtering options can be removed by simply clicking the *Remove Filters* button, bringing the matrix visualization back to its original state.

4.4. Saving Analytical States

The flexibility of filtering the rules to focus the analysis on the parts of interest allows to improve analysis scalability, but it can be easy for users to forget the rules resulting from previous different filtering options. Our implementation allows for saving the current analysis state to overcome this problem, including rules and matrix ordering. Once a state is saved, it can be either retrieved or deleted. Another small functionality is for users to export the current rules to Comma Seperated Values (CSV) files. This helps generate an offline file with rules that can be useful for subjacent tasks.
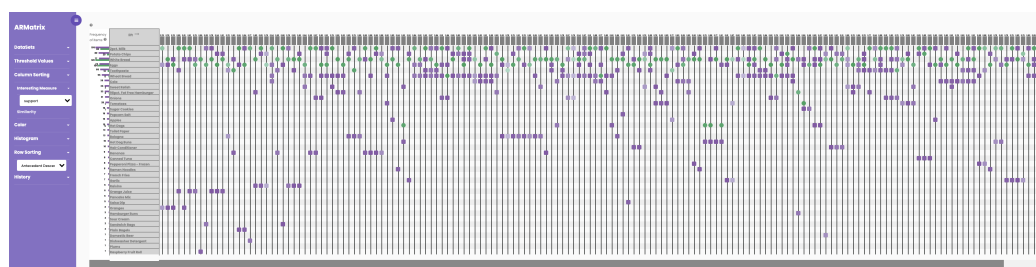
5. Use Cases and Evaluation

In this section, we present two use cases showing how to apply *ARMatrix* to analyze association rules and the results of a user evaluation, comparing our solution with the currently common practice used to analyze such rules. The default values of interesting measures for extracting the rules were manually set on a per-dataset basis and are mentioned in the respective use cases.
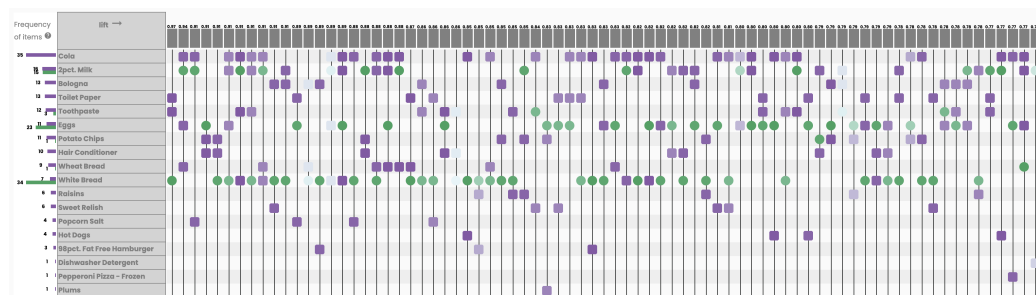
5.1. Use Cases: Market Basket Dataset

The most popular example of association rules usage is the analysis of market basket transactions. In this use case, we use the dataset from [41] consisting of 1361 transactions and 255 different items, where each transaction consists of items bought together by a client. We use typical values of support (0.02) and confidence (0.7) to extract the rules, resulting in 364 rules composed of 38 items that occur together in at least 2% of the transactions, and, in which, when an antecedent happens, a consequent also happens in at least 70% of the cases.

For this use case, we have Sam, a shopkeeper who wants to understand how to organize his convenience store and is dealing with a shortage of items. He plans to use our system to re-shelf the aisles considering space constraints and analyze the implications of out-of-stock items on the sale of other items. Sam starts his analysis by looking at all rules arranged in the item-to-rule matrix to have a general idea of the most popular items sold in his store (Figure 4a). To ease this analysis, he rearranges the rows of the matrix based on the frequency of items being antecedents and columns according to support. He does this by clicking the row and column sorting drop-down lists on the *ARMatrix* control panel, which results in a matrix where the most popular items bought in his store are displayed at the top of the item-to-rule matrix, and the more frequent rules involving the same items are placed from left to right.

(**a**) Complete set of association rules.

(**b**) Association rules containing only out-of-stock items.

Figure 4. Visualization of different sets of association rules derived from the market basket dataset. In (**a**), the complete set of rules is displayed with the matrix ordered so that the most popular items are placed at the top of the matrix with more frequent rules from left to right. This allows for the analysis of the impact of the shortage of some items (antecedents) on selling high profitable, popular items (consequents). In (**b**), only the rules containing out-of-stock items are displayed, helping to identify what products should be re-stocked first to not affect profit.

Using this visualization, he finds out that only a few items are popular among the items he displays in his store, with a large prevalence of *2pct. Milk* , *Potato Chips*, *White Bread*, *Eggs*, *Toothpaste*, and *Wheat Bread* (first 6 rows in the matrix). However, interestingly, he saw that the three rules with the highest support contain *Oranges* as an antecedent (first three columns in the matrix), potentially indicating that the sale of *2pct. Milk*, *Eggs*, and *White Bread* may depend on oranges. To further investigate what those rules convey, he adds another layer of detail, setting the color brightness to the confidence measure (not displayed here). While analyzing the results, he found that even though the rule has high support,

it presents low confidence, meaning that even though the items usually occur together, this is not frequent. However, he also discovers that other rules present similar support but much higher confidence, such as $\{WheatBread, Tomatoes\} \Rightarrow \{Eggs\}$, indicating an interesting purchase pattern in his clients' behavior that to boost eggs sale, he needs to think in strategies to boost the sale of tomatoes and wheat bread. Furthermore, more seriously, a shortage of tomatoes and wheat bread may affect the sale of highly profitable item, eggs.

This raises some red flags, and after getting a general idea of the relationships existing in his transaction dataset, he saves his current analytical state and focuses on the list of currently unavailable items. This list includes *Raisin*, *Toilet Paper*, *Plums*, *Bologna*, *Dishwasher Detergent*, *Hair Conditioner*, and *Cola*. To analyze the repercussions of missing items on the sale of other products, he selects these items in the filtering panel, selecting the rules where they appear as antecedent or consequent. After ordering the resulting matrix rows based on item antecedent frequency and columns based on lift, Sam finds out that the number of rules having *Dishwasher Detergent*, *Plums*, and *Raisins* is small (indicated by the histograms in the rows), meaning a lower effect in store's sales if these items are not re-stocked (Figure 4b). With a low budget in this period, he decides not to order these items. On the other hand, *Bologna*, *Toilet Paper*, and *Cola* (first rows in the matrix) more frequently appear in the rules, indicating that their absence may affect other products' sales, especially the highly profitable item *White Bread* since these items are common antecedents in rules where the consequent is *White Bread*. To stay inside his budget and increase sales margin, he orders more *Bologna*, *Toilet Paper*, and *Cola* quantities.

After dealing with re-stocking, he now focuses on re-shelving items. Given the limited size of his convenience store, he needs to select few items to display. He starts analyzing breakfast items by retrieving the initial analysis state containing all rules, discovering that only 6 rules contain *Pancake Mix* (a traditional breakfast item). He then filters the rules to show only these 6 rules to conduct a detailed analysis. He first orders the matrix columns by confidence, rows by item antecedent frequency, and set the color brightness to confidence as well to help understand the differences among rules from the truthfulness of such rules. Based on the resulting visualization, he discovers that the rules with highest and lowest confidences, $\{WhiteBread, PancakeMix\} \Rightarrow \{Eggs\}$ and $\{PancakeMix, WheatBread\} \Rightarrow \{WhiteBread\}$, respectively, seem to present an interesting behavior (Figure 5). By double-clicking in these rules, a radar chart visualization is shown to compare them in detail. He immediately sees that, although confidence is slightly larger for the first rule (0.906 vs. 0.848), lift of the second rule is considerably larger (0.936 vs. 0.792), indicating that the probability of people buying *White Bread* given that they are buying *Pancake Mix* and *Wheat Bread* is much more prominent in practice, so he decides to closely display these three items.
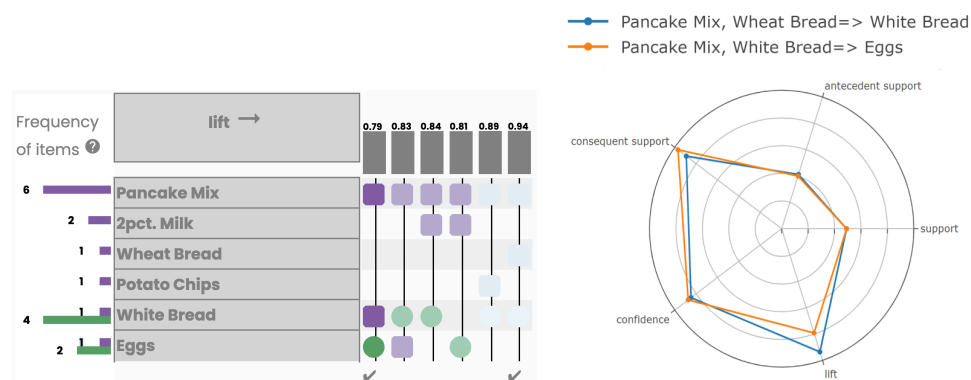

Figure 5. Visualizations of all rules containing *Pancake Mix*. The rules presenting the largest lift and good confidence indicates that usually when *Pancake Mix* and *Wheat Bread* are bought, *White Bread* is also bought with high probability. So, one possible way to improve the sell of this last item is to put it near the other two.

5.2. Use Case: Heart/Medical Dataset

In the second example, we show how to use association rules and *ARMatrix* to analyze non-transactional datasets. We use the *Statlog Heart* dataset [42] containing information about 270 patients (age, sex, chest pain, resting blood pressure, among others) and the presence or not of heart disease. To transform this dataset in a way that it can be processed using association rules analysis, the information about each patient needs to be converted into transactions. We do this by considering the unique values of each categorical attribute as items. For instance, a patient declared as male contains in his transaction the item *Gender = male*. The continuous (and integer) attributes are discretized into bins, defining categories representing potential items. The presence of absence of disease is also transformed into items.

In this use case, we have Dr. Rae, a cardiologist wanting to investigate the inter-association of symptoms and coronary diseases. After setting the initial values of 0.15 for support and 0.9 for confidence, she gets 119 rules with 17 different items including both symptoms/causes and disease. Dr. Rae decides to analyze only rules containing the disease presence or absence as consequent, so she selects *Disease = Absent* and *Disease = Present* as consequent items in the filtering panel. After filtering, she obtains 46 rules, 43 with disease absent, and 3 rules with disease present.

To construct the initial matrix visualization, Dr. Rae uses similarity to sort the columns. She notices that there is one more item, *Thal = normal*, which acts as a consequent. To distinguish it from the disease items, she sorts the rows by item antecedent frequency (Figure 6 top). Using the visualization, Dr. Rae readily sees that the three rules with 'Disease=Present' as consequent are very similar in terms of symptoms, with the presence of *Thal = reversible defect* and *Chest pain = 4.0* in the three. The only difference is that one rule also involves *Exercise = Yes* and another *RBP = 121–150*. To further investigate, she double clicks these rules for comparison in the radar chart (Figure 6 bottom). By checking all the interesting measures, she sees that they have pretty much the same values in terms of lift and confidence. The only difference is consequent support for one rule. She realizes that the low value for the rule $\{Thal = reversible\,defect, Chest\,pain = 4.0\} \Rightarrow \{Disease = Present\}$ is expected since it is a subset of the other two rules and start investigating the meaning of each symptom value to better understand the implications of this result.

Based on her investigation, she discovers that *Thal = reversible defect* indicates the heredity β-thalassemia syndrome, a blood disorder that reduces the production of hemoglobin, leading to a lack of oxygen in tissues and heart failure [43]. In addition, that *Chest pain = 4.0* indicates typical angina discomfort when the heart does not get enough blood or oxygen, usually caused by blockage or plaque buildup in the coronary arteries reducing the supply of blood to the heart muscle [44]. The interesting thing in these three rules is that, as expected, those symptoms associated with high systolic blood pressure, the *RBP = 121–150* item, leads to heart problems, but, unexpectedly, that exercise behavior, the *Exercise = Yes* item, does not help to alleviate the problem and may have a detrimental effect. She decides that this is an important finding and deserves further investigation. She shares her findings with other experts and starts looking for different datasets and sources of information to understand better the roots of this association. It is important to note that even though this finding emerged from analysis and is supported by the dataset we are using, no experts were directly consulted to check its validity. Therefore this example should be considered only for didactic reasons.
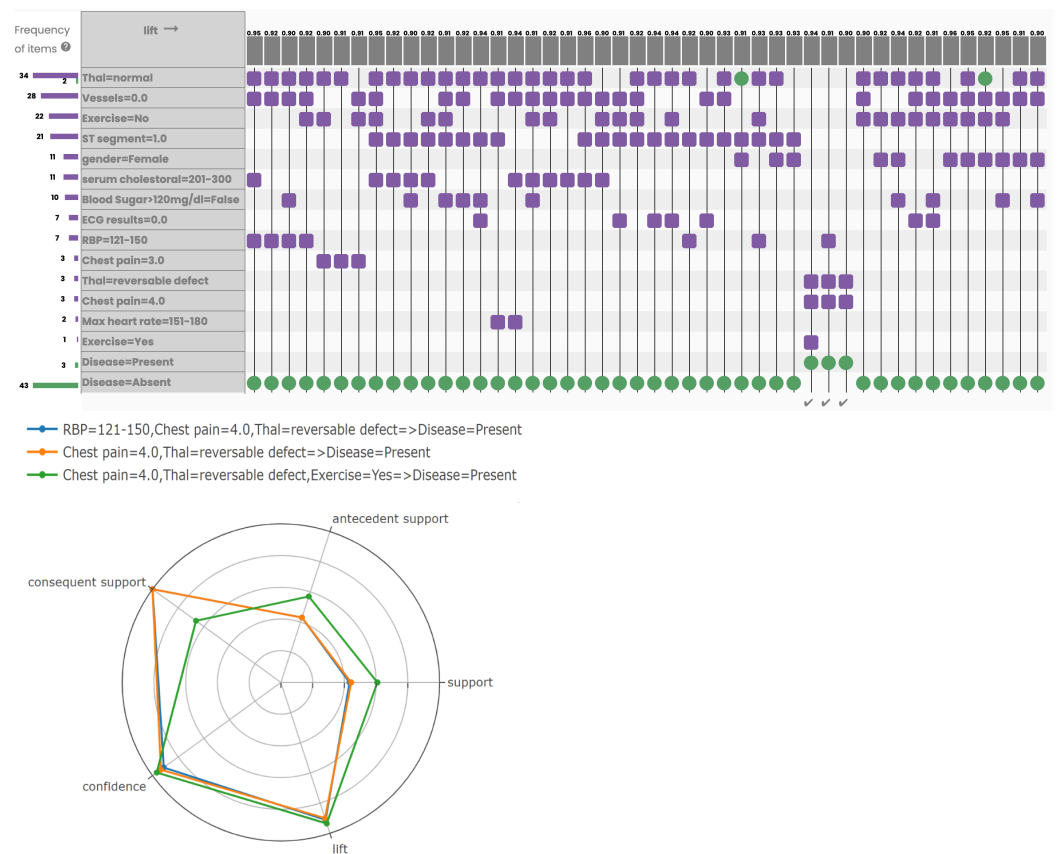
Figure 6. Visualization of rules correlating the presence or not of heart diseases and different symptoms/causes. Only three rules have as a consequent the presence of heart diseases. In these rules, the related symptoms/causes involve a hereditary blood disease, reduced blood supply to the heart, high systolic blood pressure, and, unexpectedly, exercising behavior. Since these rules have very similar high confidence and lift (have a high probability of happening), this is an important finding.

5.3. User Evaluation

We conducted a study to evaluate the effectiveness of the *ARMatrix*. We aimed to analyze the ease of use and know if the users can get meaningful insights about rules while using our system. Users answered 4 different surveys, including demographic, analytical-related, usability, and task load (NASA-TLX) questions. Table 1 present these questions. All users answer questions using an online system without supervision (due to COVID-19 restrictions).

Twenty participants were selected for the study, out of which 3 are data analysts, 1 is an inventory specialist, 6 are IT specialists (Software Developer, Data Engineer, etc.), and 4 work either in restaurants or shops. The rest of the participants are graduate and doctoral students. Twelve participants declared a good understanding of association rules, and the remaining very little knowledge. Seventeen are familiar with data analytics tools; three are neutral.

Table 1. List of the survey and questions used in the user evaluation.

Survey	Alias	Question
Demo-graphics	D1	How familiar are you with Association Rules?
	D2	How familiar are you with Data Analytics Tools, such as Microsoft Excel or Tableau?
	D3	What is your area of study/work?
	D4	What is the highest level of education you have completed?
Analysis	T1	Task: Check the system for visual rules of the market basket data.
		Question: What is the number of rules with "Tomatoes" as an antecedent.
	T2	Task: Check the excel sheet to examine the rules.
		Question: What is the number of rules with "Potato Chips" as a consequent.
	T3	Task: Filter the rules with white bread being an antecedent.
		Question: What is the frequency of Eggs as an consequent?
	T4	Task: Using the Excel sheet, filter the rules with the onion being an antecedent.
		Question: What is the frequency of Eggs as an consequent?
	T5	Task: Filter the rules with garlic being an antecedent and compare all the rules by double-clicking items in the column.
		Question: Which rule has both high support and high confidence?
	T6	Task: Using the Excel sheet, filter the rules with pancake mix being an antecedent.
		Question: Which rule have high confidence, leverage and conviction?
	T7	From the above questions, which approach's answers are you confident about?
	T8	From the above questions, which approach is more efficient?
Usability	U1	The functionalities are easy to use.
	U2	The rules were easily understood without prior knowledge about the datasets
	U3	I think that I would like to use this system frequently
	U4	I found the system unnecessarily complex
	U5	I thought the system was easy to use
	U6	I found the various visualizations in this system were well integrated
	U7	I found the system very cumbersome to use
	U8	I think that I would need the support of a technical person to be able to use this system
	U9	I thought there was too much inconsistency in this system
Task Load	L1	On a scale of 1-10, how mentally demanding task was it to use the system?
	L2	On a scale of 1–10, how physically demanding task was it to use the system?
	L3	On a scale of 1–10, how hard you had to work to answer the questions using the system?
	L4	On a scale of 1–10, how frustrated were you to answer the questions using the system?

After the demographic questions (*D1–D4*), each participant was provided with a 15 min video tutorial with an overview of association rules (https://youtu.be/b0camSVo0 10, 16 March 2021) and how to use the *ARMatrix* framework (https://youtu.be/7wvWPZYZF7I, 18 March 2021). They were asked a set of quantitative and qualitative questions (*T1–T8*) followed by two subjective questions to get overall feedback (*T9–T10*). For the study, we used the Market Basket dataset (see Section 5.1), and the users were asked to interact with the rules using the *ARMatrix* visual system and the textual rules in the form of a spreadsheet, the usual format employed by commercial solutions [12]. For the analysis using the spreadsheet, participants could use any software of their choice, where each row includes a lists of antecedents and consequents and all expected association rule metrics, such as support, confidence, lift, leverage and conviction. Figure 7 exemplifies the difference between ARMatrix and a possible spreadsheet viewer application.
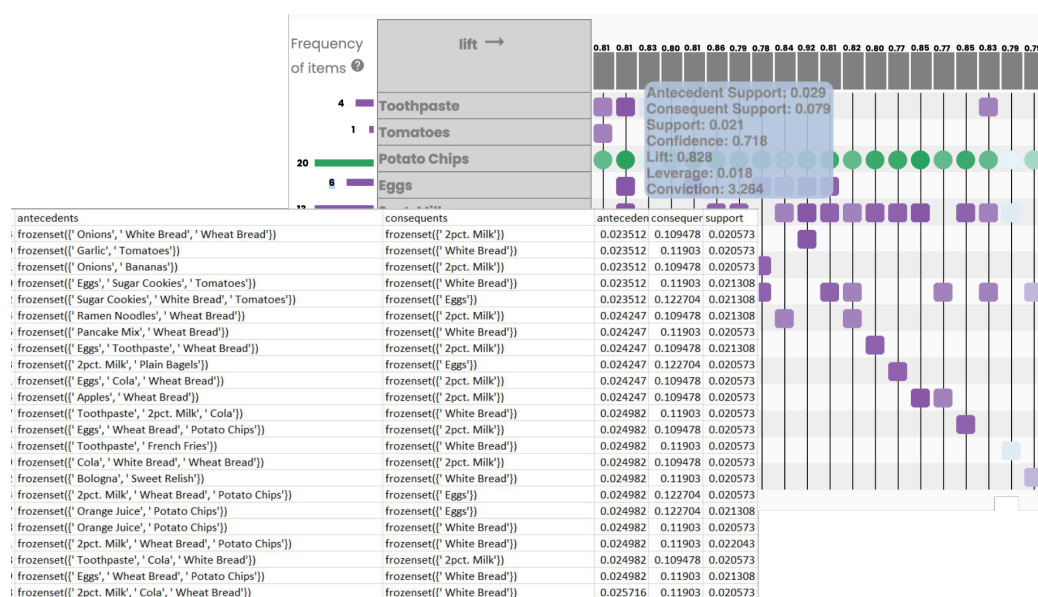
Figure 7. Example comparison of ARMatrix and the spreadsheet view utilizing Microsoft Excel.

The average time taken by the participants to complete all the tasks along with qualitative and subjective questions was 32′10″. We executed the test alternating questions' order and the employed tool (spreadsheet or *ARMatrix*). The results are summarized in Table 2. Using *ARMatrix*, users found the correct answer 95% of the time (the mistake was always committed by the same user, but we decide not to remove it since we do not have evidence that this was intentional). Since the tasks got progressively more complex, the spreadsheet participant's accuracy lowered as opposed to ARMatrix participants, as we can see in Table 2.

Given that participants performed tasks with *ARMatrix* and spreadsheets, we also use *paired t-test* [45] to analyse correlated tasks following Table 2. For this test, we give 1 to correct answers and 0 to wrong ones and assume as null-hypothesis (H_0) that the average of answers using *ARMatrix* or a spreadsheet is the same. For the first pair of tasks (*T1 % T2*), the $\rho = 0.16 > 0.05$, so the H_0 cannot be discarded, indicating that *ARMatrix* is not statistically better than a spreadsheet in this case. However, for the second (*T3 % T4*) and the third pairs (*T5 % T6*) of tasks, $\rho = 0.01 < 0.05$ and $\rho = 0.002 < 0.05$, rejecting H_0 and supporting that *ARMatrix* can attain statistically significant better results. This is not surprising since a simple one-dimensional ordering can answer the first pair of tasks (easily supported using a spreadsheet), but the other two pair of tasks cannot, especially the last one, where the comparison needs to be pairwise. Therefore, it is possible to conclude that *ARMatrix* is better in answering analytical tasks correctly if the task is more complex.

Table 2. Comparing participants accuracy using *ARMatrix* and the usual spreadsheet for the analysis of association rules. Using *ARMatrix*, participants attained better results, especially when the task gets more complex and demand more intricate analysis to be answered.

Task	Participant's % of Correct Answers	
	ARMatrix	**Spreadsheet**
(T1–T2) What is the frequency of item i being an antecedent/consequent?	19/20 = 95%	17/20 = 85%
(T3–T4) Filter the rules with item i as antecedent and find the frequency of item j being a consequent	19/20 = 95%	13/20=65%
(T5–T6) Filter and compare the rules to find the rule with highest support and confidence	19/20 = 95%	11/20 = 55%

After the tasks, we asked users open-ended questions (*T7–T8*). All participants unanimously said they were more confident and efficient using *ARMatrix* when compared to spreadsheets. We also ask some usability questions (*U1–U9*) which, through the results presented on Figure 8, indicates that *ARMatrix*, in general, is easy to use and understand. As is expected, due to the distribution of experience in analysis tasks among the users, around 30% of participants said they would still need the support of a technical person to use *ARMatrix* in the future. The user evaluation ends with NASA-TLX questions (*L1–L4*), which assess the user's effort during the test. The average value of answers on a scale between 0 and 10 were 2.7, 1.5, 2.55 and 1.6 for tasks *L1–L4*, respectively. Overall, users found the use of *ARMatrix* not very demanding or frustrating to solve the proposed tasks.

We received further feedback from users. One of the data analysts said that the "System was easy to use and understand. The system can have many applications with the rule-based models". One participant (an inventory specialist with many years of experience) liked the idea and stated "If generalized for data for every superstore, I would definitely use it and recommend it for inventory". Among positive responses, some participants also suggested that a better way to show the different subsets can fasten the decision-making process based on these rules.
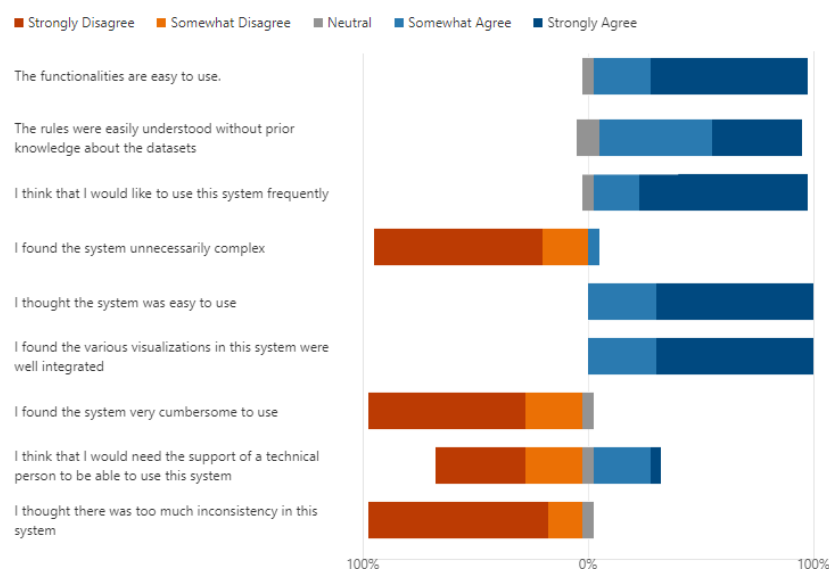

Figure 8. List of the survey sections and their questions.

6. Discussion and Limitations

In general, our three goals when designing ARMatrix were fulfilled with promising results if compared to the usual table-based approach of analyzing association rules. However, depending on what is considered large, the user's available display size remains a bottleneck. Although our method supports thousands of rules to be calculated and displayed at once, ARMatrix uses a non-overlapping matrix representation which is mainly limited by the user's display size and resolution. Through internal tests, regular FullHD laptop displays can show a few hundred rules and items without much problem before ARMatrix resorts to horizontal scroll, already rendering a considerable improvement from previous techniques. However, alternative solutions to reduce the need of horizontal scroll will need to be sought in order to display more rules in the same display. One potential solution is to aggregate rules and represent groups of rules instead of single rules, allowing for recovering details on demand. Another is to define automatic ways to set filtering parameters offering users with possibilities where relevant group patterns are found. The challenge is to define what relevant means and how to meaningfully group rules. Either way, the key concept is to offer auto-multiscale analysis, which is a complex concept and an open problem.

One issue we observed during our tests and user evaluation is that different information, sometimes conflicting, can be deduced if different perspectives are considered. This is, in fact, a problem related to the association rules mining nature of creating combinations of variables (items) without control of the meaning of the combinations. Rules are only probable casual relationships. Therefore, as in any exploratory analysis, domain knowledge is indispensable to validate the findings. For instance, the result we present in the second use case relating exercise with heart diseases can only be validated by experts and should not be considered conclusive. In this sense, having a visual representation that helps understanding rules' truthfulness based on different interesting measures and allows for quick exploration and comparison is very powerful.

Finally, despite the compelling results we obtained in our user study, some points need consideration. First, we have informally discussed our results with an expert in supermarket stocking, and, despite the encouraging feedback saying the tool may be an interesting asset, there is still work to do to make the actual use of *ARMatrix* practical. Second, our option to compare *ARMatrix* with spreadsheets can be considered, up to an extent, unfair. The reason for us selecting spreadsheets instead of other visualizations is based on three major factors: (1) spreadsheets are still be the most common representation used by commercial tools to analyze association rules [12], such as through Microsoft Excel and IBM SPSS; (2) the absence of readily-available implementations of other interactive visual metaphors to be used in a production web-based environment for a fair comparison; and (3) the inherent limitations of the existing metaphors as previously discussed. For example, graph-based approaches cannot convey "if-then" relationships, and 3D matrices need special interactive apparatus to make sense. So, any other comparison could also be considered unfair if either they are not supported by the correct devices, or cannot properly support the interactive analytical tasks, or are not source-available for comparison, as is discussed previously in the related work.

7. Conclusions and Future Work

In this paper, we present *ARMatrix*, a visualization framework for analyzing association rules based on an item-to-rule matrix visual metaphor where columns are rules and rows, items. Using *ARMatrix* users can execute different analytical tasks considering datasets and rules of reasonable sizes, allowing for the comparison of rules and set of rules considering antecedents and consequents while supporting details on demand. *ARMatrix* usefulness is demonstrated through two use-cases involving different scenarios and a user study where participants are asked to use it to solve common tasks. The results are promising. *ARMatrix* is more accurate than the usual approach to analyze association rules, and users feel more confident in their conclusions, rendering *ARMatrix* a powerful tool.

For future work, we plan to increase the visual scalability of our tool by developing a top-down approach by clustering similar rules and display rules in a cluster on demand, reducing the total number of rules to be displayed at a time. Another interesting direction of future work is to provide support to visualize the change occurring in rules when the data are updated over time in dynamic scenarios. Furthermore, the association rule visual metaphors of *ARMatrix* could potentially be adapted to other challenging applications, such as the visualization of a transactional dataset's metadata (e.g., functional dependencies and relaxed functional dependencies) within the areas of *Data Profiling* [46] and *Data Management* [47]. We reserve such investigations for future work.

Author Contributions: Conceptualization, R.V. and F.V.P.; investigation, R.V., L.C. and F.V.P.; methodology, R.V. and F.V.P.; supervision, F.V.P.; validation, R.V., L.C. and F.V.P.; visualization, R.V.; writing—original draft, R.V.; writing—review and editing, L.C. and F.V.P. All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: The study protocol was approved by the Institutional Review Board (or Ethics Committee) of Dalhousie University (REB # 2020-5406, 4 December 2020).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Tan, P.N.; Steinbach, M.; Kumar, V. *Introduction to Data Mining*, 1st ed.; Pearson: London, UK, 2005; p. 792.
2. Hilasaca, G.M.; Paulovich, F.V. Visual feature fusion and its application to support unsupervised clustering tasks. *Inf. Vis.* **2020**, *19*, 163–179. [CrossRef]
3. Agrawal, R.; Imieliński, T.; Swami, A. Mining Association Rules between Sets of Items in Large Databases. *SIGMOD Rec.* **1993**, *22*, 207–216. [CrossRef]
4. Sagin, A.; Ayvaz, B. Determination of Association Rules with Market Basket Analysis: Application in the Retail Sector. *Southeast Eur. J. Soft Comput.* **2018**, *7* .://dx.doi.org/10.21533/scjournal.v7i1.149.g136 [CrossRef]
5. Tajbakhsh, A.; Rahmati, M.; Mirzaei, A. Intrusion detection using fuzzy association rules. *Appl. Soft Comput.* **2009**, *9*, 462–469. [CrossRef]
6. Atluri, G.; Gupta, R.; Fang, G.; Pandey, G.; Steinbach, M.; Kumar, V. Association Analysis Techniques for Bioinformatics Problems. In *Bioinformatics and Computational Biology*; Rajasekaran, S., Ed.; Springer: Berlin/Heidelberg, Germany, 2009; pp. 1–13.
7. García, E.; Romero, C.; Ventura, S.; Calders, T. Drawbacks and solutions of applying association rule mining in learning management systems. *CEUR Workshop Proc.* **2007**, *305*, 13–22.
8. Vellido, A.; Martín-Guerrero, J.; Rossi, F.; Lisboa, P. Seeing is believing: The importance of visualization in real-world machine learning applications. In Proceedings of the ESANN 2011 Proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning, Bruges, Belgium, 27–29 April 2011; pp. 27–29.
9. Yamamoto, C.; Oliveira, M.C.; Rezende, S. Including the user in the knowledge discovery loop: Interactive itemset-driven rule extraction. In Proceedings of the 2008 ACM Symposium on Applied Computing (SAC), Ceará, Brazil, 16–20 March 2008; pp. 1212–1217.
10. Yamamoto, C.H.; de Oliveira, M.C.F.; Fujimoto, M.L.; Rezende, S.O. An Itemset-Driven Cluster-Oriented Approach to Extract Compact and Meaningful Sets of Association Rules. In Proceedings of the Sixth International Conference on Machine Learning and Applications (ICMLA 2007), Cincinnati, OH, USA, 13–15 December 2007; pp. 87–92.
11. Chunsheng, Z.; Yan, L. The visual mining method of Apriori association rule based on natural language. In Proceedings of the 2016 7th IEEE International Conference on Software Engineering and Service Science (ICSESS), Beijing, China, 26–28 August 2016; pp. 572–575.
12. Castillo, W.; Rojas, A.P.; Meneses, C. Augmented Visualization of Association Rules for Data Mining. In Proceedings of the 8th Alberto Mendelzon Workshop on Foundations of Data Management, Cartagena de Indias, Colombia, 4–6 June 2014.
13. Yamada, S.; Funayama, T.; Yamamoto, Y. Visualization of relations of stores by using Association Rule mining. In Proceedings of the 2015 13th International Conference on ICT and Knowledge Engineering (ICT Knowledge Engineering 2015), Bangkok, Thailand, 18–20 November 2015; pp. 11–14.
14. Yamamoto, C.H.; Cristina, M.; Oliveira, F. Visualization to Assist Users in Association Rule Mining Tasks. Ph.D. Dissertation, University of São Paulo, São Paulo, Brazil, 2005.
15. Klemettinen, M.; Mannila, H.; Moen, P.; Toivonen, H.; Verkamo, A. Finding Interesting Rules from Large Sets of Discovered Association Rules. In Proceedings of the Third International Conference on Information and Knowledge Management, Gaithersburg, MD, USA, 29 November–1 December 1994; pp. 401–407.
16. Chen, W.; Xie, C.; Shang, P.; Peng, Q. Visual analysis of user-driven association rule mining. *J. Vis. Lang. Comput.* **2017**, *42*, 76–85. [CrossRef]
17. Hahsler, M.; Chelluboina, S. Visualizing Association Rules in Hierarchical Groups. In Proceedings of the 42nd Symposium on the Interface: Statistical, Machine Learning, and Visualization Algorithms, Cary, NC, USA, 1–3 June 2011.
18. Couturier, O.; Hamrouni, T.; Yahia, S.B.; Nguifo, E.M. A scalable association rule visualization towards displaying large amounts of knowledge. In Proceedings of the 2007 11th International Conference Information Visualization (IV '07), Zurich, Switzerland, 4–6 July 2007; pp. 657–663.
19. Said, Z.B.; Guillet, F.; Richard, P.; Picarougne, F.; Blanchard, J. Visualisation of Association Rules Based on a Molecular Representation. In Proceedings of the 2013 17th International Conference on Information Visualisation, London, UK, 16–18 July 2013; pp. 577–581.
20. Ounifi, M.S.; Amdouni, H.; Elhoussine, R.B.; Slimane, H. New 3D Visualization and Validation Tool for Displaying Association Rules and Their Associated Classifiers. In Proceedings of the 2016 20th International Conference Information Visualisation (IV), Lisbon, Portugal, 19–22 July 2016; pp. 152–158.

21. Blanchard, J.; Guillet, F.; Briand, H. Exploratory Visualization for Association Rule Rummaging. In Proceedings of the KDD-03 Workshop on Multimedia Data Mining, Seattle, WA, USA, 22–25 August 2004.

22. Szafir, D.A. The Good, the Bad, and the Biased: Five Ways Visualizations Can Mislead (and How to Fix Them). *Interactions* **2018**, *25*, 26–33. [CrossRef]

23. Ankerst, M. Visual Data Mining and pixel-oriented visualization techniques. In Proceedings of the SIGKDD Workshop on Visual Data Mining, San Francisco, CA, USA, 26–29 August 2001; pp. 85–88.

24. Fernandez-Basso, C.; Ruiz, M.; Calvo-Flores, M.; Martin-Bautista, M. 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, Prague, Czech Republic, 9–13 September 2019.

25. Ong, K.h.; Ong, K.l.; Ng, W.K.; Lim, E.P. CrystalClear: Active Visualization of Association Rules. In Proceedings of the International Workshop on Active Mining, in conjunction with the IEEE International Conference on Data Mining (ICDM), Melbourne, FL, USA, 19–22 November 2003.

26. Fister, I. Association rules over time. In *Frontiers in Nature-Inspired Industrial Optimization*; Springer: Berlin/Heidelberg, Germany, 2022; pp. 1–16.

27. Menin, A.; Cadorel, L.; Tettamanzi, A.; Giboin, A.; Gandon, F.; Winckler, M. ARViz: Interactive Visualization of Association Rules for RDF Data Exploration. In Proceedings of the 2021 25th International Conference Information Visualisation (IV), Sydney, NSW, Australia, 5–9 July 2021; pp. 13–20. doi: 10.1109/IV53921.2021.00013. [CrossRef]

28. Alyobi, M.A.; Jamjoom, A.A. A Visualization Framework for Post-Processing of Association Rule Mining. *Trans. Mach. Learn. Data Min.* **2020**, *13*, 83–99.

29. Caruccio, L.; Deufemia, V.; Polese, G. Visualization of (multimedia) dependencies from big data. *Multimed. Tools Appl.* **2019**, *78*, 33151–33167. doi: 10.1007/s11042-019-07951-0. [CrossRef]

30. Wong, P.C.; Whitney, P.; Thomas, J. Visualizing association rules for text mining. In Proceedings of the 1999 IEEE Symposium on Information Visualization (InfoVis'99), San Francisco, CA, USA, 24–29 October 1999; pp. 120–123.

31. Chakravarthy, S.; Zhang, H. Visualization of Association Rules over Relational DBMSs. In *Proceedings of the 2003 ACM Symposium on Applied Computing*; Association for Computing Machinery: New York, NY, USA, 2003; SAC '03, pp. 922–926.

32. Liu, G.; Suchitra, A.; Zhang, H.; Feng, M.; Ng, S.K.; Wong, L. AssocExplorer: An association rule visualization system for exploratory data analysis. In Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Beijing, China, 12–16 August 2012.

33. Xu, Q.; Li, C.; Xiao, B.; Guo, J. A visualization algorithm for alarm association mining. In Proceedings of the 2009 IEEE International Conference on Network Infrastructure and Digital Content, Beijing, China, 6–8 November 2009; pp. 326–330.

34. Heike, H.; Siebes, A.; Wilhelm, A. Visualizing association rules with interactive mosaic plots. In Proceedings of the sixth ACM SIGKDD international conference on Knowledge discovery and data mining, Boston, MA, USA, 20–23 August 2000; pp. 227–235.

35. Slimani, T.; Lazzez, A. Efficient Analysis of Pattern and Association Rule Mining Approaches. *Int. J. Inf. Technol. Comput. Sci.* **2014**, *6*, 70–81. [CrossRef]

36. Heravi, M.; Zaïane, O. A study on interestingness measures for associative classifiers. In Proceedings of the ACM Symposium on Applied Computing, Sierre, Switzerland, 22–26 March 2010; pp. 1039–1046.

37. Brin, S.; Motwani, R.; Ullman, J.; Tsur, S. Dynamic itemset counting and implication rules for market basket data. *ACM SIGMOD Rec.* **2001**, *26*, 255–264. [CrossRef]

38. McNicholas, P.; Murphy, T.; O'Regan, M. Standardising the lift of an association rule. *Comput. Stat. Data Anal.* **2008**, *52*, 4712–4721. [CrossRef]

39. Raschka, S. MLxtend: Providing machine learning and data science utilities and extensions to Python's scientific computing stack. *J. Open Source Softw.* **2018**, *3*, 638. [CrossRef]

40. Michael, B.; Ogievetsky Vadim, H.J. D3: Data-Driven Documents. *IEEE Trans. Vis. Comput. Graph.* **1997**, *7*, 2–3.

41. Market Basket Dataset. 2021. Available online: http://csci.viu.ca/barskym/teaching/DM2012/labs/LAB7/PartII.html (accessed on 10 May 2020).

42. Dua, D.; Graff, C. Statlog (Heart Dataset), UCI Machine Learning Repository. 2017. Available online: https://archive.ics.uci.edu/ml/datasets/statlog+(heart) (accessed on 10 May 2020).

43. Kremastinos, D.T.; Farmakis, D.; Aessopos, A.; Hahalis, G.; Hamodraka, E.; Tsiapras, D.; Keren, A. Beta-thalassemia cardiomyopathy: History, present considerations, and future perspectives. *Circ. Heart Fail.* **2010**, *3*, 451–458. [CrossRef] [PubMed]

44. Hombach, V.; Höher, M.; Kochs, M.; Eggeling, T.; Schmidt, A.; Höpp, H.W.; Hilger, H.H. Pathophysiology of unstable angina pectoris—Correlations with coronary angioscopic imaging. *Eur. Heart J.* **1988**, *9*, 40–45. [CrossRef] [PubMed]

45. Kim, T. T test as a parametric statistic. *Korean J. Anesthesiol.* **2015**, *68*, 540. [CrossRef]

46. Naumann, F. Data profiling revisited. *ACM SIGMOD Rec.* **2014**, *42*, 40–49. [CrossRef]

47. Fekete, J.D. Visual analytics infrastructures: From data management to exploration. *Computer* **2013**, *46*, 22–29. [CrossRef]