InterVisAR: An Interactive Visualization for Association Rule Search

Chih-Wen Cheng
Dept. of Biomedical Engineering
Georgia Institute of Technology and
Emory University
Atlanta, GA, USA
cwcheng83@gatech.edu

Ying Sha
School of Biology
Georgia Institute of Technology
Atlanta, GA, USA
ysha8@gatech.edu

May D. Wang
Dept. of Biomedical Engineering
Georgia Institute of Technology and
Emory University
Atlanta, GA, USA
maywang@bme.gatech.edu

ABSTRACT
Association rule mining has been utilized extensively in many areas because it has the ability to discover relationships among variables in large databases. However, one main drawback of association rule mining is that it attempts to generate a large number of rules and does not guarantee that the rules are meaningful in the real world. Many visualization techniques have been proposed for association rules. These techniques were designed to provide a global overview of all rules so as to identify the most meaningful rules. However, using these visualization techniques to search for specific rules becomes challenging especially when the volume of rules is extremely large. In this study, we have developed an interactive association rule visualization technique, called InterVisAR, specifically designed for effective rule search. We conducted a user study with 24 participants, and the results demonstrated that InterVisAR provides an efficient and accurate visualization solution. We also verified that InterVisAR satisfies a non-factorial property that should be guaranteed in performing rule search. All participants also expressed high preference towards InterVisAR as it provides a more comfortable and pleasing visualization in association rule search comparing with table-based rule search.

Categories and Subject Descriptors
J.3 [Applied Computing]: Life and medical sciences – health and medical information system.

General Term
Algorithm

Keywords
Association rules, knowledge visualization, visualization techniques, methodologies.

1. INTRODUCTION
As information technology advances, the exponential growth and availability of data is leading us from conventional small data to the era of “big data.” However, the biggest challenge in the era is the complexity of transforming the data into useful knowledge to make better decisions and actions.

Association rule mining (ARM) is a data mining method that discovers meaningful relationships among variables in databases. Agrawal et al. first introduced the concept of ARM to extract regularities between products in large-scale transaction databases, as known as the market basket analysis [1]. Association rules are in the form of X⇒Y, which means that X implies Y, where X and Y are called antecedent and consequent, respectively. In its original marketing analysis context, a rule X⇒Y carries the meaning that if a customer purchases items X, he/she is also likely to purchase items Y.

Two basic metrics —support and confidence— quantify the frequency and the level of association of a rule. The support of an association rule is defined as

\[ \text{Supp}(X \Rightarrow Y) = \frac{\text{count}(X \cup Y)}{N}, \]

where count(X∪Y) indicates the number of data tuples that contain both X and Y, and N is the total number of data tuples in the database. Support metric measures the portion of the database that contains the rule, implying the rule’s frequency. The other metric, confidence, is defined as

\[ \text{Conf}(X \Rightarrow Y) = \frac{\text{count}(X \cup Y)}{\text{count}(X)}. \]

Confidence metric calculates the ratio of data tuples that contain both X and Y to the records that contain only X. On the other hand, the confidence reveals the level of association between X and Y.

The goal of ARM is to find all frequent and confident association rules that satisfy two user-specified thresholds: minimum support (\(\text{Suppmin}\)) and minimum confidence (\(\text{Confmin}\)). The main advantage of ARM is its ability to discover a huge quantity of rules covering all possible relationships that reside in the database. Generally, applying ARM on a database with \(n\) distinct items can generate a set of rules in the order of \(O(2^n)\). However, this fact also becomes one of its main drawbacks. It is because the mining process does not guarantee that all the rules found are meaningful in the real world, even though they are frequent and confident [2, 3]. Therefore, we need post-hoc processes to improve the quality of
rules and identify more interpretable, meaningful, and actionable rules.

Most of the ARM post-hoc process efforts can be categorized into two directions. Efforts in the first direction try to improve the quality of rules by pruning redundant rules such as by applying significant tests [4] and identifying non-actionable rules [5]. Although these methods can dramatically remove meaningless rules, the resulting rules are still often too large for human interpretation. Therefore, the second post-hoc process direction is to apply visualization approaches to transform rules into graphical forms for human’s natural perception.

Most of rule visualization techniques are designed to provide a global overview or a summary of all rules so as to highlight and identify the top confident (or frequent) rules. However, a user may want to search a specific rule even though it is not highlighted. For example, we may have already highlighted a meaningful rule \( \{ \text{soda}, \text{chips} \} \Rightarrow \{ \text{burger} \} \). But we cannot apply this rule to a customer who purchases soda and potatoes but no chips. We need to search for another specific rule \( \{ \text{soda}, \text{beer} \} \Rightarrow \{ \text{burger} \} \) to customize this transaction. Therefore, there is a need for visualization techniques allowing effective search for a specific rule even though this rule does not have top confidence (or support) value. However, none of the current rule visualization techniques were originally designed for rule search, using them to rummage a specific rule is inefficient and not intuitive.

In this paper, we propose a novel visualization technique, called InterVisAR, designed to provide efficient rule search. The visualization is expected to be helpful for researchers, informaticians, data scientists, and data analytics who tend to discover knowledge in large-scaled association rule base. Our goal is to improve three searching aspects: efficiency, accuracy, and the overall accurate rules per minute. In Section 2, we first make a brief review of visualization techniques in association rule mining and describe InterVisAR in Section 3. The user study is a key step to evaluate a new visualization technique [6, 7]. Therefore, in Section 4, we report and discuss a user study conducted to quantitatively evaluate the following hypotheses:

- \( H1 \). In rule search task (\( T1 \)), InterVisAR outperforms table-based rules in all three aspects.
- \( H2 \). In rule search task (\( T1 \)), InterVisAR satisfies the non-factorial property in all three aspects.
- \( H3 \). In next-level item search task (\( T2 \)), InterVisAR outperforms table-based rules in all three aspects.
- \( H4 \). In next-level item search task (\( T2 \)), InterVisAR satisfies the non-factorial property in all three aspects.

Regarding to the user preference, we also design our user study to evaluate the following two hypotheses:

- \( H5 \). InterVisAR is preferred in rule search (\( T1 \)).
- \( H6 \). InterVisAR is preferred in next-level item search (\( T2 \)).

To the best of our knowledge, this is the first quantitative user evaluation study for association rule visualization. Finally, we conclude this study in Section 5.

2. RELATED WORKS AND PROBLEMS

2.1 Related Works

Visualization analytics is the process facilitated by interactive visual and graphical interfaces so as to reduce the load on working memory, offloads cognition, and harnesses the power of human perception [8]. In healthcare, visualization has been used to enhance the delivery of knowledge mined from complex and large clinical datasets. However, these visualization techniques were only designed to provide an overview of data or knowledge after analysis. It is still challenging to apply these visualization techniques to search for patient-specific knowledge.

The scatter plot is a straight-forward 2-D visualization of association rules [9]. A scatter plot uses support and confidence on the two axes so that each rule is represented as a dot. The color of a dot can display one more rule metric with a color map on the side of the plot (e.g., lift metric [10] in Fig. 1-a). A variation of scatter plot is called the two-key plot in which the color of a dot indicates the number of items in the rule [11]. Bayardo and Agrawal suggest that the most interesting rules can be easily highlighted using scatter plot since they are usually located on the

\[ \text{Figure 1. Three examples of conventional association rule visualization. (a) Scatter plot with 3,000 association rules. The color indicates the lift metric [10]. (b) Matrix-based visualization with the same 3,000 rules in (a). The color indicates the confidence metric. (c) Graphic-based visualization of nine rules from three items (i.e., Soda, Bread, and Ham). The width of the connection represents a rule’s confidence value.} \]
Matrix-based visualization is another popular technique for association rules. The x- and y-axis represent all possible itemsets in antecedents and consequents. Due to the limited space of axis labels, itemsets are displayed in numbers to shorten long itemsets. An association rule is plotted at the intersection of its antecedent and consequent itemsets. In the 2-D visualization, the color of the intersection can be used for a selected metric (e.g., confidence in Fig. 1-b). An extension version can use 3-D bars where the height of a bar represents the second rule metric [12].

Graphic-based visualization [13] represents rules using vertices and edges with directions. A rule is a connection starting from an antecedent vertex to another consequent vertex. The width of edges can be used to indicate a selected rule metric (e.g., confidence in Fig. 1-c).

2.2 Current Challenges and Motivations

Because none of the aforementioned visualization techniques was originally designed for rule search, it is difficult to search a specific rule from a large number of rules. In this section, we discuss some limitations in current association rule visualization techniques, which make it challenging to apply in biomedical and healthcare-related data.

In scatter plots, each rule is represented as a dot. The user can only hover the pointer over the dot to read its antecedent and consequent. It makes it very inefficient to search for a specific rule that is not located on the border, especially when the total number of rules is voluminous.

Matrix-based visualization techniques typically show numerical itemset IDs on the x- and y-axis due to the limited axis space. Rules with numerical IDs are not human-readable and require a mapping process to translate the IDs to real item names. In addition, it is very common that a large number of possible itemsets are available, making the plot either very condensed or extend very long along the x- and y-axis. Due to these facts, performing rule search using matrix-based visualization becomes non-intuitive and laborious.

Graphic-based visualization can only handle a very small number of rules because it becomes cluttered as the result of increased number of vertex and edges. Therefore, graphic-based visualization is intrinsically not suitable for searching specific rules.

2.2.1 Non-factorial Properties

Assuming that we have N possible antecedent items with a fixed consequent, the N items can generate up to 2^N - 1 raw rules regardless of $Supp_{min}$ and $Conf_{min}$. To search a target rule with an l-itemset (i.e., a set of l items) in antecedent, the user needs to extract a pool of rules that all have l items in antecedents (first step) and, afterwards, rummage throughout the extracted rules for the target l-itemset (second step). Such a searching process poses two main problems especially when N is large.

Firstly, the searching efficiency mainly depends on the volume of the extracted rules in the first step, which may have size up to $2^\frac{N}{l}$. For instance, when searching for a rule with a 4-item ($l = 4$) antecedent from a dataset of eight available items ($N = 8$), we need to extract and rummage the rule through a pool of $C_4^8 = 70$ rules. If we add one more available item to the database, i.e., $N = 9$, the size of the extracted rule pool increases to $C_4^9 = 126$. That is, we increase 80% of search loading by simply adding one available item even for the same target antecedent.

Secondly, it is intuitive that rules with long antecedents require longer searching time than those with short antecedents. However, in a dataset with eight available items ($N = 8$), searching for a rule with a 4-item antecedent may require a longer time compared to searching for another rule with a 7-item antecedent, which counters our intuition. That is because the former requires the user to rummage in a pool of $C_4^8 = 70$ rules, but only $C_7^8 = 8$ rules in the later.

Based on the two aforementioned challenges, we can summarize that the effort required to search a rule should only monotonically increase for longer antecedents, instead of being factorially influenced by the number of total possible items N and the targeted antecedent length l (i.e., $C_l^N$). We call this non-factorial property and addressing this property is one of our main motivations.

3. InterVisAR

3.1 Components and Features

InterVisAR is a post-hoc process performed after all rules are generated by ARM. It is based on an assumption that all rules share the same consequent specified during the mining process. It is not necessary to prune rules a priori, meaning that the visualization can handle all possible rules regardless of the $Supp_{min}$ and $Conf_{min}$.

As depicted in Fig. 2, the visualization consists of rows of horizontal lines. Y-axis represents full names of the row items, instead of showing their ID numbers (e.g., in the matrix-based visualization). X-axis, spanning from 0% to 100%, is used for support and confidence values. Two static vertical dashed lines show the $Supp_{min}$ (in green) and $Conf_{min}$ (in blue). Each row is composed of one horizontal support line (in green) and one horizontal confidence line (in blue). The visualization updates these row items iteratively based on user selections.

Initially, without any selection, the support and confidence lines of each item span from 0% to the potential support and confidence values, e.g., Fig 2-a. If any item has been selected (i.e., the antecedent itemset is not empty), the plot adds two vertical solid lines displaying the current support, $Supp_{current}$, in green and confidence, $Conf_{current}$, in blue, e.g., Fig 2-b. Afterwards, each remaining item is updated accordingly; the support line spans from $Supp_{current}$ to one end indicating the change of support value if we add the potential row item to the current selected antecedent itemset. Similarly, the confidence line spans from $Conf_{current}$ to one end indicating the potential change of confidence value. A solid dot located at the end indicates that the change is significant; otherwise, a hollow dot is presented. These graphic components provide the following usages:

1. We can determine which potential items can significantly increase/decrease the current support, i.e., towards the right/left of the vertical $Supp_{current}$ line. Similarly, we can determine which potential items can significantly increase/decrease the current confidence, i.e., towards the right/left of the vertical $Conf_{current}$ line.

2. We can determine which potential row items can still keep the rule frequent, i.e., the horizontal support (green) lines remain on the right of vertical $Supp_{min}$ line. Similarly, we can tell which potential items can still keep the rule confident in comparison with the vertical $Conf_{min}$ line.

3. We can determine if the rule of current selected antecedent itemset is frequent, i.e., the vertical $Supp_{current}$ line is on the top border of the scatter [9].
right of the vertical $\text{Supp}_{\text{min}}$ line. Similarly, we can determine if the rule of current selected antecedent itemset is confident, i.e., the vertical $\text{Conf}_{\text{current}}$ line is on the right of the vertical $\text{Conf}_{\text{min}}$ line.

Several interactive features are available to make the visualization friendlier. When the user hovers the cursor over a row item (e.g., ‘Low SpO2’ in Fig 2-a), the visualization displays corresponding values at the ends of the horizontal lines. In addition, upon selection (by mouse click) of a row item, the visualization updates the “Selected Items” table by adding the newly selected item (e.g., ‘Low SpO2’ in Fig 2-b).

### 3.2 Algorithm

The InterVisAR algorithm updates the visualization components based on interactive user selections. We describe the setting of the algorithm as follows. Assuming we have extracted a set of raw rules $R$ in which each rule $r$ has four elements, including $r.a$ as the antecedent item-set, $r.l$ as the length (i.e., the number of items of $r.a$), $r.sup$ as the support value, and $r.conf$ as the confidence value. The goal of the process is to find a rule with a target antecedent itemset. $\text{Supp}_{\text{min}}$ and $\text{Conf}_{\text{min}}$ are also known algorithm inputs.

The visualization starts by displaying all rules with 1-item antecedents. The process keeps updating the visualization by interactively receiving the user’s selection of next-level items. Each update refreshes the remaining next-level row items, horizontal support and confidence lines, and the vertical $\text{Supp}_{\text{current}}$ and $\text{Supp}_{\text{current}}$ lines. The level-by-level process continues until the rule with target antecedent itemset has been found. A reset function is provided to restart the entire process.

The complete algorithm of InterVisAR is provided in Fig. 3. We provide some notes about the algorithm:

- Line 1 draws $\text{Supp}_{\text{min}}$ and $\text{Conf}_{\text{min}}$ in two vertical dashed green and blue lines.
- According to the $\text{sort\_type}$ input, line 2 sorts raw rules $R$ in one of the three ways: alphabetically (a to Z or Z to a) by row item names or numerically (descending or ascending) by potential support or confidence values. Fig. 2-a and 2-b are two examples in which row items are sorted descendingly by confidence values. Fig. 4 is another example of alphabetical sorting.
- Line 3 clears $r_{\text{current}}$ by setting $r_{\text{current}.a} = \emptyset$, $r_{\text{current}.l} = 0$, $r_{\text{current}.sup} = 0$, and $r_{\text{current}.conf} = 0$.
- Line 4 updates the two vertical $\text{Supp}_{\text{min}}$ and $\text{Conf}_{\text{min}}$ lines to $r_{\text{current}.sup}$ and $r_{\text{current}.conf}$, respectively.
- In line 5, the $\text{extractNextLevelRules}(r_{\text{current}}, R)$ extracts a rule pool $W$ in which each next-level rule $r$ satisfies $r.l = r_{\text{current}.l} + 1$ and $r_{\text{current}.a} \subseteq r.a$.  

**Figure 2.** An example of InterVisAR visualization algorithm with seven items. (a) The initial plot without any selected items. The cursor hovers on a next-level items “Low SpO2” with its corresponding potential support, confidence, and maximal confidence. (b) The updated plot after the item “Low SpO2” has been selected. Three out of the six potential items can change the confidence significantly (two increases and one decrease).
Algorithm InterVisAR($R$, $Supp_{min}$, $Conf_{min}$, $sort_{type}$)

Input: $R$ is the extracted rule set, $Supp_{min}$ is the minimum support, $Conf_{min}$ is the minimum confidence, and $sort_{type}$ is the type of sorting

1. Draw $Supp_{min}$ and $Conf_{min}$;
2. Sort $R$ by $sort_{type}$;
3. Clear $r_{current}$;
4. Update $Supp_{current}$ & $Conf_{current}$ to $r_{current}.conf$ & $r_{current}. supp$;
5. $W = \text{extractNextLevelRules}(r_{current}, R)$;
6. while $W$ is not empty do
7. Clear the current plot;
8. for each rule $r$ in $W$ do
9. Exclude $r_{current}.a$ from $r.a$ and label it in y-axis;
10. Add a support bar from $r_{current}.supp$ to $r.supp$;
11. Add a confidence bar from $r_{current}.conf$ to $r.conf$;
12. $(\text{sig supp}, \text{sig conf}) = \text{BinomTest}(r_{current}, R)$;
13. if $\text{sig supp}$ then Plot solid circle at $r.supp$;
14. else Plot hollow circle at $r.supp$;
15. if $\text{sig conf}$ then Plot solid circle at $r.conf$;
16. else Plot hollow circle at $r.conf$;
17. while an item $i$ is selected from $r_{current}.a$ do
18. Exclude $i$ from $r_{current}.a$ goto line 4;
19. while an rule $r$ is selected do
20. $r_{current} = r$ goto line 4;
21. while reset; goto line 3;
22. while reset; goto line 3;

Figure 3. The InterVisAR algorithm.

- If $W$ is not empty in line 6, the algorithm clears the plot in line 7. Otherwise, the visualization halts in line 22 and waits for the user to reset.
- From line 8 to line 16, each next-level rule $r$ in $W$ is processed one-by-one.
- Line 9 excludes $r_{current}.a$ from each $r.a$ and labels the remaining item in y-axis.
- Line 10 adds a horizontal support (green) line from $r_{current}.supp$ to $r.supp$, and line 11 adds a horizontal confidence (blue) line from $r_{current}.conf$ to $r.conf$.
- Lines 12 tests if $r_{current}.supp$ and $r.supp$, and $r_{current}.conf$ and $r.conf$, are statistically different. We use binomial test since the outcome of association rules is binary (i.e., the consequent is hold or not). The test is performed at two-tailed 95% significant level. If the change of support is significant, line 13 places a solid green circle at the end of the support line; otherwise, places a hollow circle (line 14). Similarly, if the change of confidence is significant, line 15 places a solid blue circle at the end of the confidence line; otherwise, places a hollow circle (line 16).
- After updating all next-level rules in $W$, the system halts and waits for one out of three user actions that can trigger line 17, 19, or 21, respectively.
- Line 17 is triggered if the user selects any item $i$ from the “Selected Items” table. Line 18 excludes the item $i$ from $r_{current}.a$, and the process repeats from line 4.
- Line 19 is triggered when a next-level rule $r$ is selected from $W$. Line 20 sets $r_{current} = r$ and repeats from line 4.
- The entire visualization can be reset anytime in line 21. If triggered, the process repeats from line 3.
- Finally, the process reaches line 22 only when the extracted $W \neq \emptyset$. Users can reset the process to line 3.

4. USER STUDY

4.1 Motivation

In one of our previous studies, we reported a system called Predictive Health Association Rule Mining (PHARM) [14]. PHARM was designed to generate quantitative and objective rules in predictive health settings [15] by leveraging a dataset from 696 healthy subjects. However, we used tables to represent rules in the PHARM system, which has been reported inefficient for searching specific rules. Thus, in this study, we use the predictive health data to evaluate the improvement of rule search using InterVisAR compared to table-based representations. Note that we did not compare InterVisAR to other rule visualization techniques because none of them was originally designed for rule-searching purposes so that the comparison may be biased. Our goal was (1) to evaluate InterVisAR’s performance in terms of accuracy and efficiency, (2) to assess if InterVisAR satisfies the non-factorial property that has been mentioned in Section 2.2.1, and (3) to demonstrate that InterVisAR is a preferred rule-searching method according to user satisfaction.

4.2 Method

4.2.1 Dataset

We used the data from the case study in [14] that discovered predictive rules for development of mental illness. The prediction
Table 1. Antecedent items and their corresponding scales

<table>
<thead>
<tr>
<th>Item</th>
<th>Scale Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>BDI=9.1</td>
<td>Beck Depression Inventory</td>
</tr>
<tr>
<td>ESSI&lt;23.9</td>
<td>Enriched Social Support Inventory</td>
</tr>
<tr>
<td>EPWORTH&gt;10.3</td>
<td>Epworth Sleepiness Scale</td>
</tr>
<tr>
<td>FACT&lt;=60.6</td>
<td>Functional Assessment of Cancer Therapy</td>
</tr>
<tr>
<td>FAD&gt;=2.3</td>
<td>Family Assessment Device</td>
</tr>
<tr>
<td>GAD7&gt; 6.1</td>
<td>Generalized Anxiety Disorder 7-item</td>
</tr>
<tr>
<td>PSSE&gt;25.4</td>
<td>Perceived Empathic Self-Efficacy Scale</td>
</tr>
</tbody>
</table>

was based on scale scores of seven existing psychological disorders, including family dysfunctioning, lack of social support, depressive, perceived empathic self-infficacy, sleepiness, dysfunctioning due to cancer therapy, and anxiety. In other words, there were totally seven available antecedent items (i.e., \( N = 7 \)) and one fixed consequent. As listed in Table 1, each antecedent item was composed of a scale name and a numerical range for disorder. The seven antecedent items allowed us to generated \( 2^7 - 1 = 127 \) raw rules that had lengths between 1 and 7. Several examples are provided in Table 2.

4.2.2 Participants

We recruited 24 participants (12 females and 12 males). All participants had normal computer operation skills. Each participant performed the experiment on an individual PC with two 20.1” wide aspect LCD monitors. The left screen displayed training information and experiment tasks, and the right one was used to perform tasks. The interface of the system was implemented in MATLAB (MathWorks, Natick, MA).

4.2.3 Tasks

Participants were asked to perform two types of tasks:

- **T1. Rule Search.** Given a target antecedent itemset, the participant tried to find the rule with that itemset and report its support and confidence values.

- **T2. Next-level Item Search.** Given a target antecedent itemset, the participant tried to find one next-level item from the remaining antecedent items. The new rule combining the given itemset with this new next-level item had the highest confidence among all remaining items.

4.2.4 Rule representations

Each task was performed by four table-based (starting with \( T \) in Table 3) and two visualization-based (starting with \( Vis \) in Table 3) rule presentations, which are:

- **T.** Table with rules sorted by ascending antecedent length.
- **TO.** In addition to **T,** in each rule, items were sorted alphabetically (a to Z) by name.
- **TL.** Table with a slider bar controlling the length of displayed antecedents. Displayed rules were sorted by descending confidence value.

**TLO.** Table with a slider bar controlling the length of displayed antecedents. Displayed rules were sorted alphabetically (a to Z) by first item name.

**VisC.** InterVisAR with next-level items sorted by descending confidence value.

**VisA.** InterVisAR with next-level items sorted alphabetically (a to Z) by name.

4.2.5 Procedure

Before each study, the participant received a training session to understand the basic concept of association rules, the PHARM dataset, the tasks, and the rule presentations. Afterwards, participants were randomly assigned to two groups (\( G1 \) and \( G2 \)) balanced by gender. As shown in Table 3, participants in two groups were associated with two different sets of rule presentations. Each set consisted of three rule presentations that are exclusive with those in the other group. Performing tasks using only three presentations, instead of all six, allowed us to control a study to be completed within 60 minutes so as to prevent fatigue from prolonged operation time. In this way, each participant performed six (2x3 task-presentation) sections. Participants were required to practice before the start of each section until they were fully ready.

Target rules in \( T1 \) (\( T2 \)) sections were associated with antecedent lengths between one and six (four). Each length was queried twice so that 12 participants generated 24 (12x2) records associated with a length. In summary, this user study was conducted as a design of 2 groups x 12 participants x 2 tasks x 3 rule presentations x 6 different lengths in \( T1 \) (or 4 in \( T2 \)) x 2 queries.

Finally, after completing a task using three different presentations, participants provided their ranking to each of the three presentations via a Likert scale from 3 (the favorite) to 1 (the worst). Multiple presentations with the same rank were acceptable.

Table 2. Examples of Antecedent Examples with Different Lengths

<table>
<thead>
<tr>
<th>Length</th>
<th>Example of Antecedent</th>
<th># of Rules of Length</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>GAD7&gt;=6.1</td>
<td>7</td>
</tr>
<tr>
<td>2</td>
<td>ESSI&lt;=23.9 &amp; PSSE&gt;=25.4</td>
<td>21</td>
</tr>
<tr>
<td>3</td>
<td>BDI&gt;=9.1 &amp; FAD&lt;=2.3 &amp; GAD7&gt;=6.1</td>
<td>35</td>
</tr>
<tr>
<td>4</td>
<td>BDI&gt;=9.1 &amp; EPWORTH&gt;=10.3 &amp; FAD&gt;=2.3 &amp; GAD7&gt;=6.1</td>
<td>35</td>
</tr>
<tr>
<td>5</td>
<td>EPWORTH&gt;=10.3 &amp; FACT&lt;=60.6 &amp; FAD&gt;=2.3 &amp; GAD7&gt;=6.1 &amp; PSSE&gt;=25.4</td>
<td>21</td>
</tr>
<tr>
<td>6</td>
<td>BDI&gt;=9.1 &amp; EPWORTH&gt;=10.3 &amp; ESSI&lt;=23.9 &amp; FAD&gt;=2.3 &amp; GAD7&gt;=6.1 &amp; PSSE&gt;=25.4</td>
<td>7</td>
</tr>
<tr>
<td>7</td>
<td>BDI&gt;=9.1 &amp; EPWORTH&gt;=10.3 &amp; ESSI&lt;=23.9 &amp; FACT&lt;=60.6 &amp; FAD&gt;=2.3 &amp; GAD7&gt;=6.1 &amp; PSSE&gt;=25.4</td>
<td>1</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>127</strong></td>
<td><strong>127</strong></td>
</tr>
</tbody>
</table>
4.2.6 Measures
The practice of evaluating visualization has been studied for decades [16]. According to Lam et al., the evaluation of visualization can be grouped into seven scenarios, including evaluation communication thought visualization (CTV), and visual data analysis and reasoning (VDAR) [17]. The effectiveness measures of our user evaluation study belong to these two scenarios.

The record of a query contained two measures, the completion time (i.e., the time from being queried to the time of answer) for efficiency and a hit or miss (i.e., if the answer is correct or not) for accuracy. The efficiency of an antecedent length in a task-presentation section was analyzed using all 24 completion time records collected from the group’s 12 participants. Similarly, the accuracy of an antecedent length in a task-presentation section was analyzed using all 24 hit/miss records.

Using the efficiency and accuracy measures, we introduced a new measure to determine a participant’s overall accurate rules per minute (ARPM) for a specific length in each task-presentation section, which is calculated as:

\[ \text{ARPM} = \frac{60}{\text{completion time (in sec)}} \times \text{accuracy}. \]

Finally, using collected results, we performed repeated-measure analysis of variance (RM-ANOVA) to investigate if performances were significantly affected by different lengths in different rule presentations.

4.3 Results and Discussion
4.3.1 Rules Search
In Section 1, we state that the goal of InterVisAR is to improve rule search performance and user satisfaction. Our user study was designed to demonstrate the goals by evaluating the study results in six hypotheses (H1 to H6). The results of rule search in three aspects, including efficiency (i.e., the inverse of the completion time), accuracy (% of hit), and ARPM, are illustrated in Fig. 5. We verified H1 by observing that InterVisAR, mainly, outperformed all table-based presentations in all aspects regardless of length, table control, and sorting types. The efficiency of InterVisAR decreased in longer rules, but the accuracy was consistently high. As for sorting types, InterVisAR with alphabetical sorting (VisA) slightly outperformed it with confidence sorting (VisC) in all three

![Figure 5. Results of rule search using six rule presentations in six different lengths: (a) completion time, (b) accuracy, and (c) accurate rules per minute (ARPM). (d) Comparisons among the six methods and their statistics of differences.](image-url)
aspects. It was reasonable because rules were searched by names instead of by confidence values.

InterVisAR also demonstrated the non-factorial property in all three aspects (H2). For example, the ARPM of InterVisAR monotonically decreased in longer rules, instead of being factorially affected by the total number of available items. On the contrary, table-based rule presentations all lacked the property with respect to all three aspects. For example, tabulated rules with length control and alphabetical sorting (TLO) lacked the property even though it was the one performed most closely to InterVisAR. We can verify it as the completion time for 6-item rules was shorter than 5-item rules.

4.3.2 Next-Level Item Search

It can be expected that the effort required is higher in the task of next-level item-search than it in the rule search task. This is because the participants need to first extract all possible next-level rules, compare confidences among them, and finally identify the one with the highest confidence. This procedure was time-consuming and tedious using table-based presentations. As expected, according to Fig. 6, InterVisAR outperformed all table-based presentations in all three aspects, verifying H3. Similar to rule search, InterVisAR with alphabetical sorting (VisA) still slightly outperformed InterVisAR with confidence sorting (VisC) except for 1-item antecedents.

The results of InterVisAR also demonstrated the non-factorial property in all three aspects, which verified the hypothesis H4. For example, the APRM using InterVisAR monotonically decreased, but not factorially, in longer antecedents. Not surprisingly, table-based presentations (other than TL) all lacked the property. It was interesting that tabulated rules controlled by antecedent length and sorted by confidence (TLO) also showed the non-factorial property. It was mainly because the participants can easily extract potential next-level rules by using the slide bar. Afterwards, the top one of the extracted rules was the final rule because they were sorted by descending confidence value, which skipped the need for rule comparison. However, InterVisAR still outperformed TLO in all aspects in addition to its non-factorial property.

4.3.3 User Preference

The results of user preference are depicted in Fig. 7. Participants were more comfortable and pleased using InterVisAR in both rule-searching (H5) and next-level item-searching tasks (H6).
compared to all other table-based presentations \( (p < 0.01) \). As for table-based presentations, participants preferred those with more rule length controls and sorting. For example, the two presentations with length control (i.e., TL and TLO) were more favored than \( T \) and \( TO \) \( (p < 0.01) \), respectively.

### 4.4 Limitation and Possible Solutions

The aforementioned results positively supported our six hypotheses. However, this study still revealed several limitations that need further considerations:

First, InterVisAR was designed based on the assumption that all association rules have a predetermined and fixed consequent. It makes the visualization more suitable for classification-based association rules in which there is one and only one target consequent (i.e., the class) [18]. However, in real-world association rule search, consequents are not necessarily and nearly fixed. Therefore, more sophisticated approaches should be proposed and investigated for flexible consequents. They should also satisfy the non-factorial property to guarantee the performance.

Second, the predictive health dataset used in our user study only consisted of seven items. We believe that the demonstration of the non-factorial property using InterVisAR would be more prominent when we significantly increase the number of all available items, which accords more with real-world cases.

Third, in the user study, we divided our 24 participants into two groups to perform tasks using two different sets of rule presentations. However, we did not have chance to cross compare the user preference between these two groups. For example, it was difficult to determine which type of sorting in InterVisAR (i.e., \( VisA \) and \( VisC \)) was preferred because no participant performed tasks using both of them. We remark on this because a few participants in \( G1 \) suggested methods being tested in \( G2 \), and vice versa.

Fourth, InterVisAR has the ability to visualize potentially maximal confidences (i.e., the T-shape locations in the plot) of all remaining next-level items. However, searching for potentially maximal confidences is extremely difficult using tabulated rule presentations. This is why we excluded this task in our user study because it might cause prolonged effort for the participants. A separate user study can be conducted for this task, but our current results have already implied that InterVisAR can outperform tabulated methods in maximal-confidence search.

Finally, the current visualization only demonstrated the capability of two basic rule metrics: support and confidence. However, rules can be measured using other objective interestingness metrics, such as conviction [19], leverage [20], interest factor [21], and \( p \)-values using the chi-square test [4]. Future versions of InterVisAR will support comprehensive interestingness metrics so that users can select metrics of interest to visualize.

### 5. CONCLUSION AND FUTURE WORK

We introduced an interactive association rule visualization called InterVisAR. InterVisAR was designed specifically for association rule searching, which is different from conventional rule visualization techniques that only provide an overview or a summary of all rules. The visualization interactively receives user inputs step-by-step until all items in the target rule have been found. Based on the user’s selection, InterVisAR visualizes information about all next-level items, including potential next-level supports and confidences, and potential maximal confidences.

We also discussed a non-factorial property that a rule-searching method should satisfy to guarantee the performance. To verify it in InterVisAR, we conducted a user study by comparing to table-based rule presentations with different rule controls and sortings. The results confirmed that InterVisAR not only outperformed table-based rule presentations (in terms of efficiency, accurate, and accurate rule per minute (ARPM)), but also satisfied the non-factorial property. Participants also expressed strong preferences for InterVisAR because it provides a comfortable and pleasing rule-searching method.

In addition to the challenges and possible solutions that we mentioned in Section 4.4, we will consider the following directions for further improvement so as to maximize InterVisAR’s usability and viability. First, as InterVisAR delivers most association rule information in a small view without overwhelming user intensive panning and scrolling, we will demonstrate its usability on small screens such as on smartphones. We believe, by utilizing the touch screens, we can provide a more interactive rule visualization with minimal restriction from display size. Second, we will evaluate InterVisAR in different real-world dataset so as to validate its viability in comprehensive data-mining settings.

### 6. ACKNOWLEDGMENTS

The authors are grateful to Dr. Chanchala Kaddi and Po-Yen Wu for their valuable comments and suggestions. We also like to thank for Dr. Greg Martin for providing the data for the user study. This research has been supported by grants from NIH (U54CA119338, 1RC2CA148265, and R01CA163256), Georgia Cancer Coalition Award to Prof. MD Wang, Hewlett Packard, and Microsoft Research.

### 7. REFERENCES


