# Enhancement of Association Rules Interpretability using Generalization

**Safaa O. Al-Mamory**
*College of Business Informatics, University of Information Technology and Communications, Iraq*
salmamory@uoitc.edu.iq
**Zahraa Najim Abdullah**
*Department of computer science, University of Karbala, Iraq*
zahraa_alanbary@itnet.uobabylon.edu.iq

## Abstract

Data mining has a number of common methods. One of such methods is the association rules mining. While association rules mining often produces huge number of rules, it prevents the analyst from finding interesting rules and consequently, this method is a waste of time. Visualization is one of the methods to solve such problems. However, most of the association rule visualization techniques are suffering from viewing huge number of rules. This paper provides a modification on the techniques of the visualization to help the analyst to interpret the association rules by grouping the large number of rules using a modified Attribute Oriented Induction algorithm, then; these grouped rules are visualized using a grouped graph method. Experimental results show that the proposed technique produces excellent compression ratio.

**Keywords**: Data mining, Association rules, Visualization, AOI, Grouped graph.

## الخلاصة

تعدين البيانات يمتلك عدد من الطرق الشائعة، وان تعدين قواعد الاقتران هي احدى تلك الطرق. بما ان تعدين قواعد الاقتران ينتج كمية هائلة من القواعد فأنه يمنع المحلل من ايجاد القواعد المهمة، وبالتالي فان هذه العملية تعتبر تضييع للوقت. العرض هو احدى الطرق لحل هكذا مشاكل. وبما أن معظم تقنيات عرض قواعد الاقتران تعاني من عرض عدد هائل من القواعد، فأن هذا البحث يوفر تعديل لتقنيات العرض لتساعد المحلل لتفسير قواعد الاقتران بواسطة تجميع عدد كبير من القواعد باستخدام خوارزمية الاستقراء الموجه للصفات المعدلة و من ثم عرض تلك القواعد المجمعة باستخدام طريقة الرسم البياني المجمع , وقد أوضحت النتائج ان التقنية المقترحة تنتج نسبة ضغط ممتازة.

**الكلمات المفتاحية:** تعدين البيانات، قواعد الاقتران، العرض، AOI، مخطط مجمع.

## 1. Introduction

Data mining has a number of common methods. Association rules (AR) are one of the classic association rule mining algorithm that aim to discover hidden relation in large data sets and analysis relations between sets of itemsets. AR are "if/then" statements, such as $X \rightarrow Y$ where X and Y are sets of elements (or items) in the dataset; given the elements of right hand side of the rule (Y, the conclusion or antecedent),given the left hand side elements of the rules(X, the conditions or consequent) (Fernandes and García 2012).

Apriori is one of the association rules algorithms that produce a large number of rules that make the analyst consume time and face difficulty in discovering the interest rules from large rules. Therefore, we used the visualization technique which is considered as one of the methods for understanding, analysis and selection of the interesting rules from the huge number of rules. In this paper, graph-based visualization technique has been proposed which has the ability to display small and understandable rules. Firstly, the rules are grouped together, then, redundant rules are omitted, this is performed by Attribute Oriented Induction (AOI) technique and finally, it will be visualized by a graph-based technique called grouped graph method. In this paper, we proposed a modification for AOI to produce new algorithm called

Modified AOI. An excellent ratio of the compressed rules and graph's vertex and edge is obtained from combining the Modified AOI and the grouped graph method.

This paper contains into six sections. So far, there is an introduction. In section two, a survey of the literature related to the subject is given. In section three, we introduce a preliminary of the method. In section four, we present how rule sets could be grouped by the new modified AOI and then visualize the rules in new grouped graph visualization technique. In section five, results are discussed, while the conclusions are given in section six.

## 2. Related Works

Our project goes into two subjects; therefore, we divided the related work into two groups. The first group is the visualization techniques. Scatter plot visualization technique uses support and confidence measures for axes and lift measure for point shading (Jr and Agrawa 1999) , while two-key plot uses the order measure for point shading (Hofmann, Unwin, and Bernt 2001).

Double decker is used for displaying one rule (Hofmann H, Siebes A 2000). Parallel coordinate uses the items and its position in the rules for axes and arrow for the rules(Yang 2003). Then Matrix-based visualization technique uses antecedent and consequent for axes and interest measures for colored rectangle (Ong et al. 2002), while (matrix3D) uses the 3D bar instead of colored rectangle. (Hahsler and Chelluboina 2011b) proposed grouped matrix-based visualization technique to enhance matrix-based by grouping the antecedent of the rules. Other techniques like Graph-based visualization technique uses vertex for items or item sets and edges for relationships (Klemettinen *et al*. 1994;Ertek and Demiriz 2006;Rainsford and Roddick 2000).

The second group is the clustering of association rules techniques. In this context (Gupta, Strehl, and Ghosh 1999) proposed a new measure that takes the distance between Association rules based on a conditional probability estimate, as in Eq. (1)

$$d_{i,j} = P\left(\overline{BS_i} \vee \overline{BS_j} | BS_i \vee BS_j\right) \qquad \textbf{(1)}$$

Where the set BS is the union of items in the left and right hand sides of rule *i*, we call $d_{i,j}$ the Conditional Market-Basket Probability (CMPB) distance and CMPB measure is used by Agglomerative Chain Clustering algorithm to find the clusters. In this algorithm, a point is joined to its closest neighbor that is found from the distance matrix. Then, all the collection of points are displayed in graphs, all points that are joined as a graph will have the same label and after that the labels of all the points will be returned. The property of the algorithm is that it scales the cluster size depending upon the density of the points in the neighborhood. At the end of the algorithm, we get a tree structure that describes the multiple levels of the clustering.

( Klemettinen *et al*., 1995; Hätönen ,1995) pruned the rules by extracting a subset that is called a rule covers from the original set of rules ,a method for reducing the number of rules by eliminating of redundancy is applied. If there are two rules, the second is more specific than the first one then the second rule is removed as redundant, so he takes from the original rule set the more general rules. The rules in the cover can be ordered and grouped to make them more understandable.

(Lent *et. al*.,1997) introduced a clustered association rule as a rule that is formed by combining similar “adjacent” association rules to form few general rules instead of a set of (attribute = value) equalities, for clustered rules he had a set of value ranges using inequalities and he considered clustered association rules as in Eq. (2) The association rule is clustered in a two-dimensional space, where each axis represents one attribute on antecedent (LHS) and the consequent (RHS) that satisfies our segmentation criteria.

$$A_1 \wedge A_2 \wedge \ldots \wedge A_n \Rightarrow G \qquad (2)$$

(Berrado and Runger, 2007) found a new method for grouping the rules, where the metarules represent the relation between the discovered rules by calculating rules with one antecedent and one consequent. A graphical presentation of the metarules is defined, where each rule is represented by a node. While, each metarule connects two nodes with a directed arc such that the originated node is the antecedent of the metarule and the destination node is its consequent. Finally, with several definitions of metarule there is a pruning process for the rules.

## 3. Preliminary

Presenting rules to a user in its simple format is a difficult issue because of the huge generated number of rules from association analysis. In this section, various rules' visualization techniques are presented starting from simple techniques to advance techniques.

Starting with the simplest techniques, the association rules visualized as a scatter plot as it is displayed in Fig. 1a-c , (Jr and Agrawa 1999) proposed technique with two axes by using support and confidence measure for each axis by dividing in equal period and lift measure for shading the point that represents the rules and the lift measure is placed to the right side as Fig. 1a. scatter plot technique views in Fig. 1b with lift measure on the y-axis and support measure on x-axis while confidence measure for shading the point, so the rules are distributed on axes according to them values of the measures and their color is increased according to the measure value of the rule.

( Hofmann *et. al.*,2001) proposed two-key plot that represents another version of a scatter plot. Unwin used the support on the x-axis, the confidence on the y-axis, and the order (number of items in rules) represents the color of the points as in Fig. 1c.

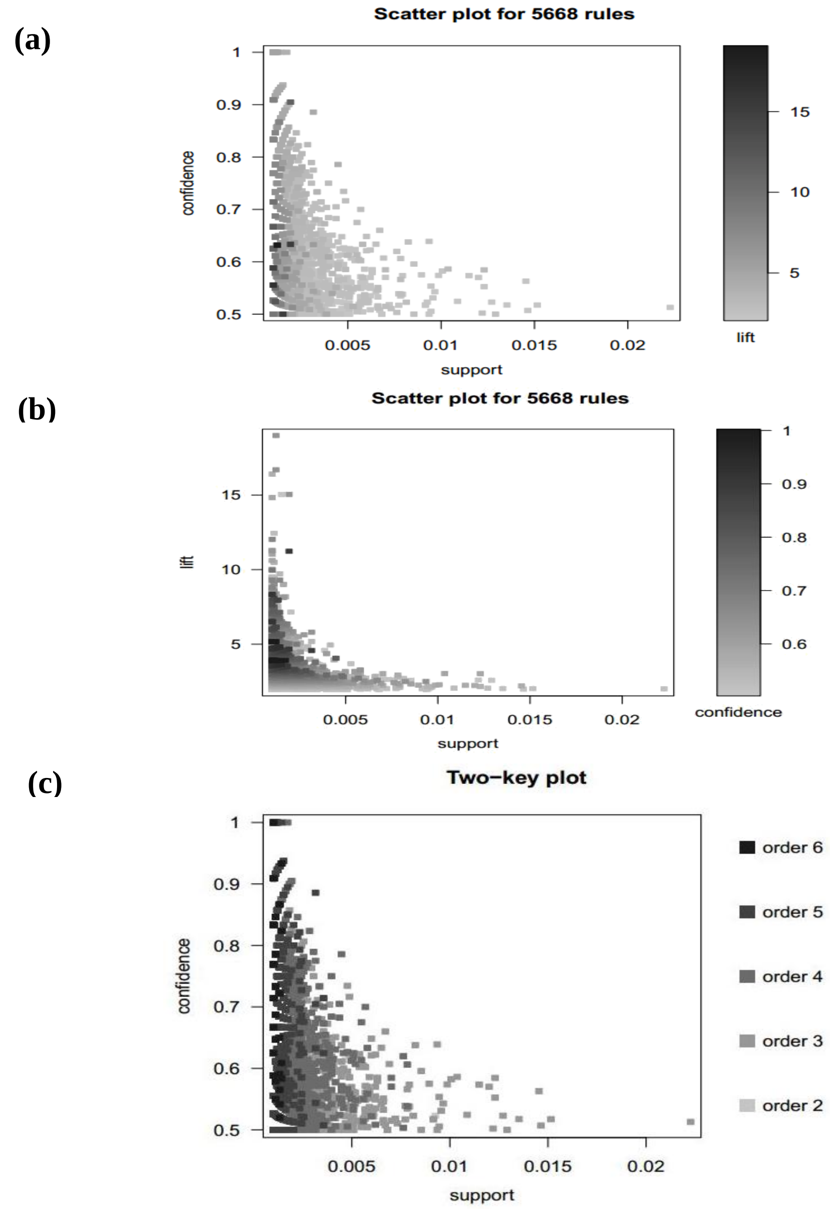


**Fig. 1** Scatter plot visualization technique with different measures and its altenative version ( tow-key plot).

Another technique of the visualization techniques is Matrix-based that uses the antecedent of the rules for x-axis and the consequent of the rules for y-axis as in Fig 2a-d. The point of intersection of x and y-axes is an interest measure of rules, if no point is drawing then there is no rules in this intersection. Formally, the visualize matrix will be constructed as in Eq. (3) that used to build this matrix with association rules as parameters for it:

$$R = \{\langle a_1, c_1, m_1\rangle, \ldots \langle a_i, c_1, m_i\rangle, \langle a_n, c_n, m_n\rangle\} \qquad (3)$$

Where *ai* represents one column of the matrix for unique LHS (antecedent) ,*ci* represents one row of the matrix for unique RHS (consequent), and *mi* represents one of the interest measures, *i* represents one rule and *n* represents the number of the rules(Hahsler and Chelluboina, 2011b), the resulting plot is shown in Fig. 2a, this plot prints several hundred of rules, then the number of row and the column of the matrix are equal to the RHS and LHS itemset, then, they reduce this rule by filtering the rule with the least confidence value. They also reordered the row and column of the matrix by similar value of interesting measure to improve such visualization as it is illustrated in Fig. 2b with two large blocks of different consequents and small for the rest of the items.

There is an alternative method for matrix-based visualization that replaced the 3D bar with the colored rectangle called (matrix3D) as in Fig. 2c that picture is displayed unordered rules but Fig. 2d is displayed order rules.

(Hahsler and Chelluboina, 2011b) introduced a method called grouped matrix-based visualization that enhances the matrix-based method by clustering the rules. Therefore, the huge number of the rules viewing in matrix based are decreased and the clustering rules will be viewed interactively by zooming in and out of a group.

They used k-means clustering for grouping the antecedent (column) into k groups, where, each group has the same consequent and the default interest measure is lift. Fig. 3 views grouped matrix using balloon with group of antecedent as row and consequent as column. (Hahsler and Chelluboina, 2011a). The top left corner is representing the larger value of lift interest measure and the lift is decreasing downwards and towards the right that produced in the sorting in the rows and the columns(Hahsler and Chelluboina, 2011a).

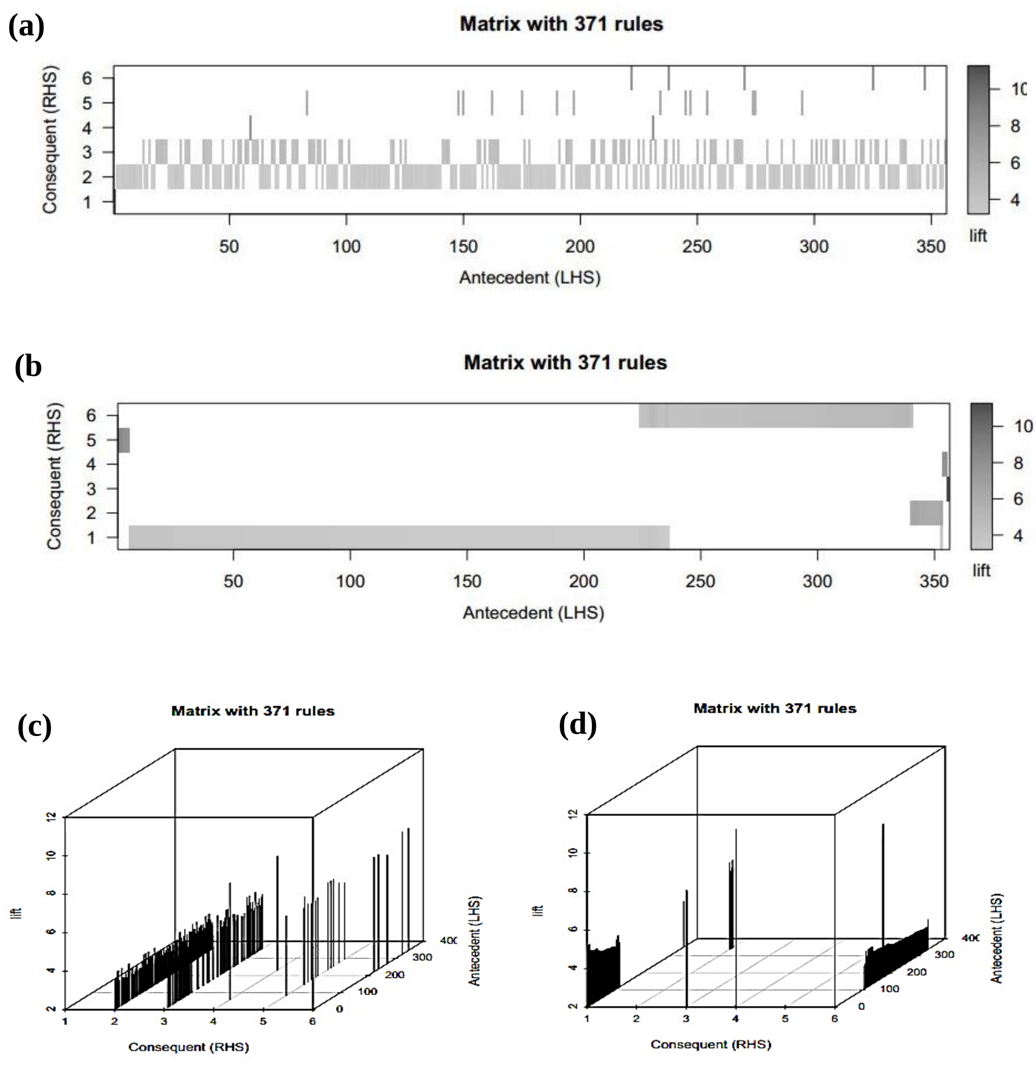


**Fig. 2** Matrix-based visualization with 3D bars (reordered and non reordered).

(Klemettinen *et al.*, 1994; Ertek and Demiriz 2006; Rainsford and Roddick, 2000) introduced graph-based method that represent the items or the itemsets by vertices and the relationship in rules by edges for visualization of association rule.

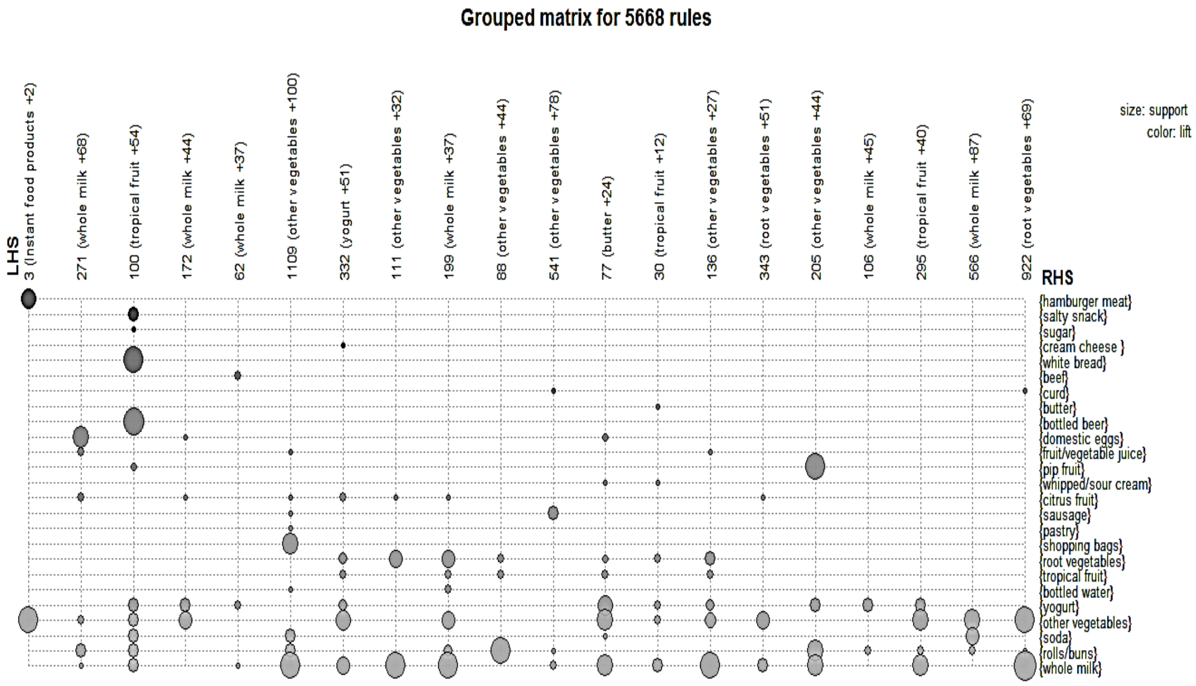


**Fig. 3** Grouped matrix-based visualization.

In Fig. 4a, the vertex is used for the itemsets and directed edges between the itemsets for the rules. In Fig. 4b, the vertex is used for the items and rules share those items. The interactive mode of this method is used either by arulesViz for small set of rules or by exploring in Gephi (Bastian *et. al.*, 2009) for large set of rules for grouping, zooming, coloring, and filtering nodes.

The multidimensional data are visualized by parallel coordinates plots, where each dimension shared on the x-axis and the y-axis separately, each data point has a line connects the values for each dimension. (Yang, 2003) used items of the nominal values on y-axis and the positions in a rule on x-axis. The head of an arrow represents the RHS; arrows go through the positions in rules on the x-axis to represent antecedent items, where shorter arrows represent the rules with less items.

Fig. 5a pictures the support for the width of the arrows and the confidence for the intensity of the color. This method prevents the displaying of the large number of the rules, crossover is the result of this method, so, to minimize the crossover of the items is ordered on y-axis as in Fig. 5b.

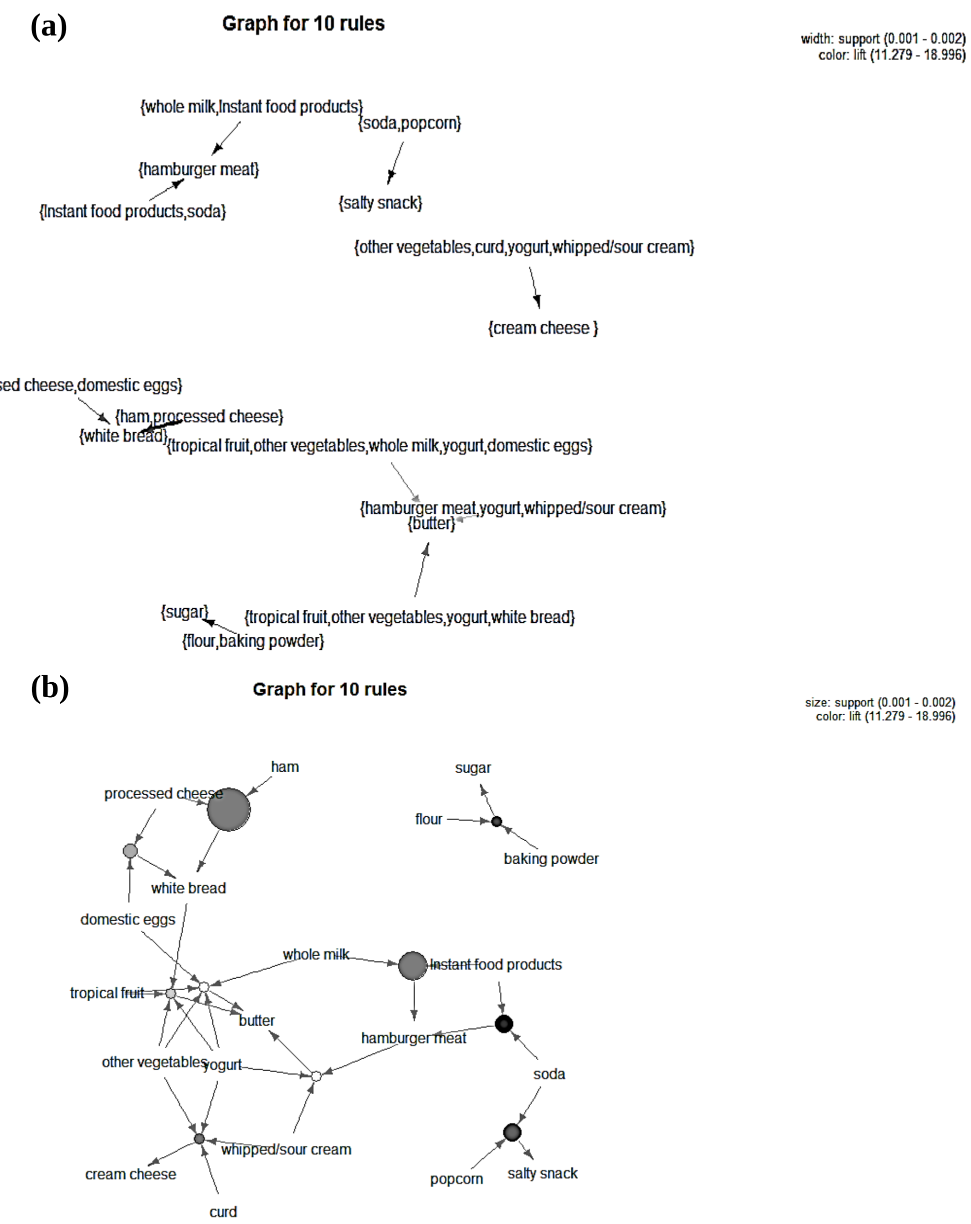


**Fig. 4** Graph-based visualization with itemsets as vertices or with items and rules as vertices.

Double decker plots use only a single horizontal split; the size of each tile is proportional to the value in the contingency table. (Hofmann *et. al.*, 2000) used double decker plots for viewing a single association rule that selecting randomly. The vertical splits are used for the items in the antecedent and horizontal highlighting is used for the consequent item. Fig. 6 shows the support that represent the area of blocks and the height of the \yes" blocks represents the confidence corresponding to the antecedent items marked as \yes."

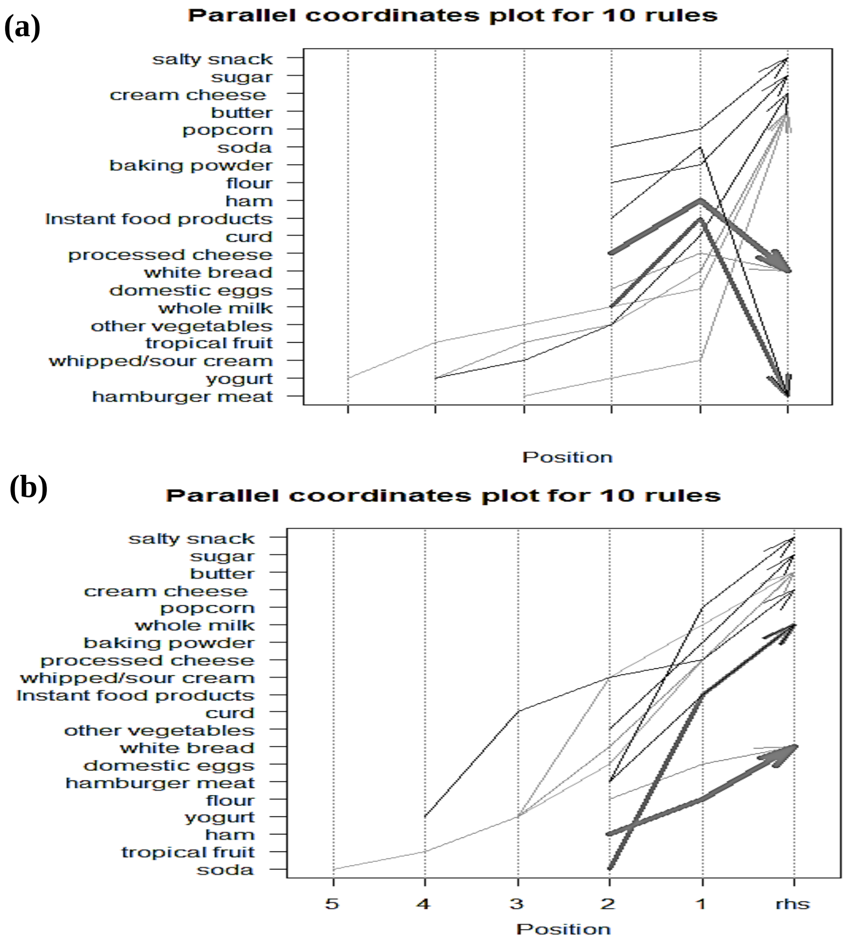


**Fig. 5** Parallel coordinate plot (with reordered and without order).

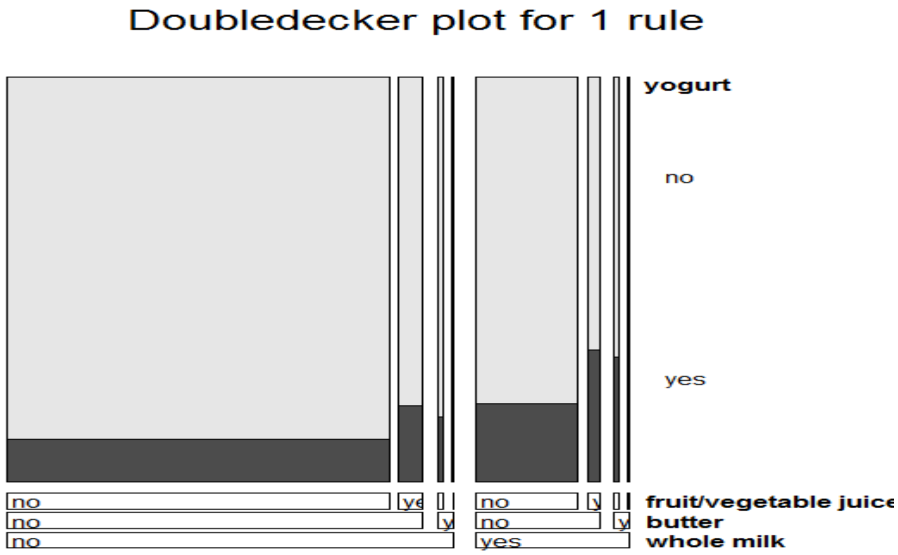


**Fig. 6** Double decker plot with one rule such as citrus fruit,butter,soda => other vegetables.

## 4. The Proposed System

Fig. 7 illustrates the proposed system flow chart, where we enhance the association rules interpretability by the following steps:

1. We take the large number of rules from the Apriori algorithm with lift, confidence and support interest measures.
2. The proposed Modified AOI algorithm is performed in Fig. 8. To produce aggregated rules that produce less number of rules than the original rules and give us overview for rules from Apriori algorithm.
3. We used one of the subjective approaches which is the visualization to determine the interesting rules. In particular, graph method used to view the aggregated rules.
4. After aggregation, the graph method called grouped graph method is applied and it represents the proposed visualization method.
5. The analyst can extract the interest rules and understand the nature of rules from the grouped graph.
6. The analyst can evaluate the system by using the measure that will be mentioned in section 5 or by other measures.

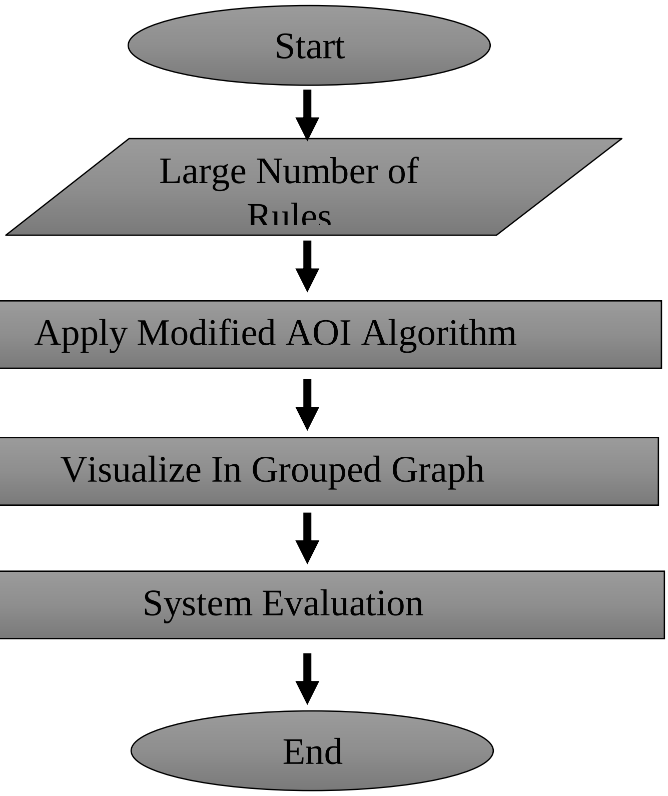


**Fig. 7** The general architecture of the proposed system.

### 4.1. Attribute Oriented Induction (AOI)

AOI technique (Hsu and Wang, 2006) is used to produce general rules or pattern from large set of rules or patterns. By two steps, attribute removal and attribute generalization perform this induction technique (Han *et. al.*, 1993). The generalization of the data by this technique is performed on the attributes depending on the concept hierarchy. The concept hierarchy represents knowledge domain. the generalization for the concepts performs from the most specific to the specific values to the root that represents single and generalized (Chen and Shen, 2005). Therefore, the AOI reduce the large data into small data by generalizing and summarizing the large data. AOI produced generalized relation from the original set of data and providing summary about the large original data (Angryk and Petry, 2005).

There are number of steps of AOI algorithm: generalization on the smallest attribute, attribute removal if there is a large set of distinct value but there is no higher-level concept, concept tree ascension, vote value should be accumulative when merging identical tuple in generalization, threshold control on each attribute that represents maximum number of distinct value of attribute in target class in the final generalized relation, generalization threshold controls on distinct tuples of generalized relation in target class, tuple is transformed to conjunction normal form and multiple tuple are transformed to disjunction normal form.

In this paper, we modified on these steps to satisfy our goal in graph visualization method and satisfy our idea as in the following algorithm. In Fig. 8, the inputs for this algorithm are hierarchy trees that are built before the generalization step that contains a number of levels that are defined before any step and a number of levels represented by Generalization Level that entered to our algorithm. In addition, the large number of rules are taken from the Apriori algorithm as input to our algorithm.

We need to reduce these rules by aggregating them, that aggregation will represent the output of this algorithm, and that is the focus of this paper. Therefore, we used the procedure of AOI algorithm for rules aggregation with a little modification, the resulted algorithm is called **Modified AOI.**

**Input**: hierarchy trees H, set of rules R, Generalization Level (GL).
**Output**: Aggregated rules.
**Method**: Modified AOI.
**Begin**
**while** GL!= 0 **do** {
  **For** each rule $R_i$ (1<=i <=n, where n=# rules) **do**
    Substitute each $itemset_k$ in the antecedent and consequent of $R_i$ by its corresponding parent in $H_k$.
  Merge identical rules
  **For** each rule $R_j$ (1<= j <n, where n=# rules) **do**
    Recomput interest measure(s) for generalized rules.
  GL=GL-1 }
**End**.

**Fig. 8** The Proposed Modified AOI Algorithm.

In step 1 the generalized level is determined and the algorithm is started from it, and in step 7 it is reduced by one in whole the algorithm until it reaches to the root of the concept hierarchy. From steps 2 and 3 , the rules generalization are made by taking every rule and are replaced each itemsets in the LHS (Left Hand Side) and in the RHS (Right Hand Side) from the rules by its corresponding parent in the hierarchy tree, then the result from these steps is a set of redundant rules, and the redundancy of the rules are removed by merging the same rules in step 4 to produce the aggregated rules. In steps 6 and 7, all the existing interesting measure of generalized rules like support, confidence and lift are recomputed that are resulted from the previous steps.

**Table 1 is illustrated the difference between traditional AOI algorithm and modified AOI by achieving or not achieving this step**

Table 1: Comparison between AOI and Modified AOI.

| Modified AOI | AOI | steps | |
|---|---|---|---|
| ✓ | ✓ | generalization on the smallest attribute | |
| × | ✓ | attribute removal if there is a large set of distinct value but there is no higher-level concept | |
| ✓ | ✓ | concept tree ascension | |
| × | ✓ | vote value should be accumulative when merging identical tuple in generalization | |
| × | ✓ | threshold control on each attribute that represents maximum number of distinct value of attribute in target class in the final generalized relation | |
| × | ✓ | generalization threshold controls on distinct tuples of generalized relation in target class | |
| × | ✓ | tuple is transformed to conjunction normal form and multiple tuple are transformed to disjunction normal form | |
| ✓ | × | Save the identical tuples before merging its that result from generalization step | |

### 4.2. Grouped Graph Visualization

In previous graph method ,small number of rules is visualized, while in this paper there is an ability to visualize large aggregated rules resulting from Modified AOI algorithm, for this reason we called this method Grouped Graph Visualization because some vertices of the rules in the graph that represents a collection of rules instead of one rule as in the previous graph method.

Grouped graph can also visualize every level in the hierarchy tree of the aggregated rules, it can visualize either drill down in the levels to show more detail about the rules or roll up in the levels to show more generalize rules that enable the user to understand large data set and take idea about the nature of the data. Example 1 illustrates this new visualization method.

**Example 1:** First four transactions are taken randomly from Groceries data set as in Table 2,then apriori algorithm are performed on these transactions to produce 13 rules as in Table 3. The Table 3 is visualized in Fig. 9a.

**Table 2: A sample of Groceries transactions**

| Items | TID |
|---|---|
| {Rolls/Buns, Pastry, Soda} | |
| {Whole Milk} | |
| {Curd Cheese, Coffee} | |
| {Red/Blush Wine, Newspapers} | |

Secondly, the aggregation is performed for rules in Table 3 by the Modified AOI algorithm with levels in Table 4 and show the result in Grouped Graph Visualization. The result from aggregation on Table 3 is 8 rules in level2 as in Table 5 and 6 rules in level1 (more generalize level) as in Table 6. Then the visualization of Table 5 is displayed in Fig. 9b and Table 6 is displayed in Fig. 9c.

**Table 3: The result 13 rules from apriori Algorithm.**

| Rules | | | Measures | | |
|---|---|---|---|---|---|
| | Left hand side | Right hand side | Support | Confidence | Lift |
| 1 | {curd cheese} | {coffee} | 0.25 | 1 | 4 |
| 2 | { coffee} | {curd cheese} | 0.25 | 1 | 4 |
| 3 | (red/blush wine} | {newspapers} | 0.25 | 1 | 4 |
| 4 | {newspapers} | {red/blush wine} | 0.25 | 1 | 4 |
| 5 | {pastry} | {soda} | 0.25 | 1 | 4 |
| 6 | {soda} | {pastry} | 0.25 | 1 | 4 |
| 7 | {pastry} | {rolls/buns} | 0.25 | 1 | 4 |
| 8 | {rolls/buns} | {pastry} | 0.25 | 1 | 4 |
| 9 | {soda} | {rolls/buns} | 0.25 | 1 | 4 |
| 10 | {rolls/buns} | {soda} | 0.25 | 1 | 4 |
| 11 | {pastry, soda} | {rolls/buns} | 0.25 | 1 | 4 |
| 12 | {rols/buns, pastry} | {soda} | 0.25 | 1 | 4 |
| 13 | {rolls/buns, soda} | {pastry} | 0.25 | 1 | 4 |

**Table 4: The levels of Groceries data set**

| Level1 | Level2 | Level3 |
|---|---|---|
| fresh products | bread and backed goods | rolls/buns |
| fresh products | bread and backed goods | Pastry |
| fresh products | dairy produce | whole milk |
| fresh products | cheese | curd cheese |
| Drinks | stimulante drinks | Coffee |
| Drinks | wine | red/blush wine |
| Drinks | non-alc. drinks | Soda |
| non-food | games/books/hobby | newspapers |

**Table 5: Aggregation 8 rules in level2 from Groceries data set containing 13 rules.**

| Rules | | | Measures | | |
|---|---|---|---|---|---|
| | Left hand side | Right hand side | Support | Confidence | Lift |
| 1 | {cheese} | Right hand side | 0.25 | 1 | 4 |
| 2 | {stimulant drinks} | {stimulant drinks} | 0.25 | 1 | 4 |
| 3 | {wine} | {cheese} | 0.25 | 1 | 4 |
| 4 | {games/books/hobby} | {games/books/hobby} | 0.25 | 1 | 4 |
| 5 | {bread and backed goods} | {wine} | 0.25 | 1 | 4 |
| 6 | {non-alc. drinks} | {non-alc. drinks} | 0.25 | 1 | 4 |
| 7 | {bread and backed goods} | {bread and backed goods} | 0.25 | 1 | 4 |
| 8 | {bread and backed goods, non-alc. drinks} | {bread and backed goods} | 0.25 | 1 | 4 |
| | | {bread and backed goods} | | | |

**Table 6: Aggregation 6 rules in level1 from Groceries data set containing 13 rules.**

| Rules | | | Measures | | |
|---|---|---|---|---|---|
| | Left hand side | Right hand side | Support | Confidence | Lift |
| 1 | {fresh products} | {drinks} | 0.50 | 0.6666667 | 0.8888889 |
| 2 | {drinks} | {fresh products} | 0.50 | 0.6666667 | 0.8888889 |
| 3 | {drinks} | {non-food} | 0.25 | 0.3333333 | 1.3333333 |
| 4 | {non-food} | {drinks} | 0.25 | 1.0000000 | 1.3333333 |
| 5 | {fresh products} | {fresh products} | 0.75 | 1.0000000 | 1.3333333 |
| 6 | {drinks, fresh products} | {fresh products} | 0.50 | 1.0000000 | 1.3333333 |

## 5. Results

At the beginning, we review the data sets that we used in the proposed system, then, we compared the visualization techniques with grouped graph and evaluated the system at the end. In Table 7, there are three data sets:

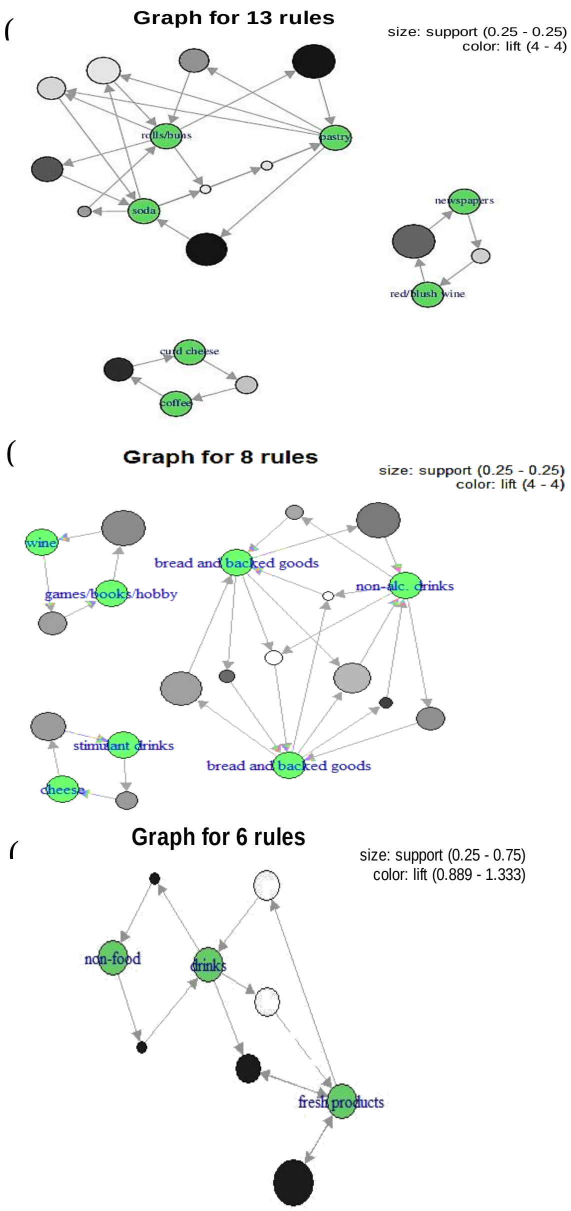


**Fig. 9** Graph visualization for original and aggregation rules in level1 and level2.

1. Groceries : is a data set with 9835 transactions and 169 items.(Hahsler, Hornik, and Reutterer 2005)
2. Adult : is a data set with 48842 observations and 115 variables.(Asuncion and Newman 2007).

Table 7 :Data set description

| Data set | Row | Column | Level3 | Level2 | Level1 |
|---|---|---|---|---|---|
| Groceries | 169 | 9835 | 169 | 55 | 10 |
| Adult | 115 | 48842 | 115 | 112 | 13 |
| Income | 50 | 6876 | 50 | 48 | 14 |

3. Income : is a data set containing 6876 observations and 50 variables.(Hastie, Tibshirani, and Friedman 2001)

In the comparison of Table 8, (Hahsler & Chelluboina, 2011b) introduced a number of the subjective measures that enable the user to select the visualization techniques according to what is needed. Grouped of rules means that this method can make grouping on rules but not the items, measures represent the number of interesting measure used by this method, interactive means that there is any type of interactive, such as zoom and select in this method and, finally, ease of use represents the ease of translation interactive Association rules using this method.

Our proposed method (grouped graph) satisfies grouped of rules more than the other visualization techniques, while other techniques are grouped items but not rules. This new technique differs from other techniques in satisfying new criteria.

Finally, Reduction Ratio is defined as a ratio of compression of some operation (Valeur 2006) as in Eq. (4) :

**Tabel 8 : Comparison between visualization methods for association rules.**

| Ease of Use* | Interactive | Measures | Grouped of Rules | Rule Set | Technique | Method |
|---|---|---|---|---|---|---|
| 2 | yes | 3 | | Large | Scatterplot | Scatterplot |
| 2 | yes | 2 + order | | Large | Two-Key plot | |
| 0 | no | 1 | | Medium | Matrix-based | Matrix |
| -2 | no | 2 | | Medium | Matrix-b. (2 measures) | |
| 1 | no | 1 | | Small | Matrix-b. (3D bar) | matrix3D |
| 0 | yes | 2 | | Large | Grouped matrix | Grouped |
| -1 | no | 1 | | Small | Parallel coordinates | Paracoord |
| -1 | no | (2) | | single rule | Double decker | Double-decker |
| 2 | no | 2 | | Small | Graph-based | Graph |
| 1 | yes | 2 | | Large | Graph-b. (external) | |
| 2 | no | 2 | √ | Large | Grouped graph | |

* it means that [-2,2] where 2:good , -2:bad

$$\text{Redction Ratio (RR)} = \frac{\text{Uncompressed number} - \text{compressed number}}{\text{Uncompressed number}} \quad (4)$$

We use this measure to be applied on the result of grouped graph method and to show the ratio of compression from the aggregation by the modified AOI technique. The result from this measure is explained in Table 9 on a number of nodes and edges that form the graph, and we show the result from the measure on the rules that resulting from aggregation of the modified AOI technique. In every data set, there is an increasing in the Reduction Ratio from level 3 down to level 1 in the number of nodes, edge, and rules by grouped graph, but there is a special case in edges of adult data set that belong to the nature of this data set because the graph method represented from edge and node is affected by the nature of the rules.

**Table 9: Reduction ratio of the proposed system on different datasets.**

| RR % | #Rules | RR % | #Edges | RR % | #Nodes | Levels | Data Set |
|---|---|---|---|---|---|---|---|
| - | 5668 | - | 22220 | - | 5778 | 3 | Groceries |
| 76 | 1335 | 78 | 4838 | 76 | 1379 | 2 | |
| 96 | 201 | 97 | 679 | 96 | 211 | 1 | |
| - | 23814 | - | 124343 | - | 23848 | 3 | Adult |
| 49.5 | 12028 | 52 | 59338 | 49 | 12059 | 2 | |
| 51 | 11738 | 48 | 64466 | 51 | 11751 | 1 | |
| - | 77781 | - | 415282 | - | 77839 | 3 | Income |
| 1 | 77288 | 0. 5 | 413356 | 0. 6 | 77343 | 2 | |
| 61 | 30163 | 58 | 174043 | 61 | 30177 | 1 | |

## 6. Conclusions

Association rules of mining algorithm generate huge rules that are difficult for analyzing and understanding, for that, we take one of the methods that helps to interpret association rules represented by the visualization method, and then we select graph visualization, but this method shows small rules. Therefore, we used AOI algorithm with a modification on it to group and reduce the large rules before visualizing them by the graph method. At the end, using the graph method with AOI algorithm produce new our technique called grouped graph visualization technique.

## 7. References


Angryk, Rafal A, and Frederick E Petry. 2005. “Discovery of Abstract Knowledge from Non-Atomic Attribute Values in Fuzzy Relational Databases.” *Elsevier*: 1–11. angryk@cs.montana.edu.

Asuncion, a, and D J Newman. 2007. “UCI Machine Learning Repository.” *University of California Irvine School of Information* 2008: 0. http:// www.ics. uci. edu/~ mlearn/MLRepository.html.

Bastian, M, S Heymann, and M Jacomy. 2009. “Gephi: An Open Source Software for Exploring and Manipulating Networks. BT - International AAAI Conference on Weblogs and Social.”

Berrado, Abdelaziz, and George C. Runger. 2007.“Using Metarules to Organize and Group Discovered Association Rules.” *Data Mining and Knowledge Discovery* 14(3): 409–31.

Chen, Yen Liang, and Ching Cheng Shen. 2005.“Mining Generalized Knowledge from Ordered Data through Attribute-Oriented Induction Techniques.” *European Journal of Operational Research* 166(1 SPEC. ISS.): 221–45.

Ertek, Gürdal, and Ayhan Demiriz. 2006. “A Framework for Visualizing Faculty of Engineering and Natural Sciences Department of Industrial Engineering A Framework for Visualizing Association Mining Results.”:593–602.m http:// research.sabanciuniv.edu.

Fernandes, Leandro a. F., and Ana Cristina Bicharra García. 2012. “Association Rule Visualization and Pruning through Response-Style Data Organization and Clustering.” *Advances in Artificial Intelligence – IBERAMIA 2012* 7637: 71–80. http://dx.doi.org/10.1007/978-3-642-34654-5_8.

Gupta, Gunjan K, Alexander Strehl, and Joydeep Ghosh. 1999. “Distance Based Clustering of Association Rules.” *Proc. ANNIE 1999, St. Louis* 9: 759–64.

Hahsler, Michael, and S Chelluboina. 2011a. “Visualizing Association Rules: Introduction to the R-Extension Package arulesViz.” *R project module*. http://www.comp.nus.edu.sg/~zhanghao/project/visualization/[2010]arulesViz.p.

Hahsler, Michael, and Sudheer Chelluboina. 2011b. “Visualizing Association Rules in Hierarchical Groups.” *42nd Symposium on the Interface: ...* (Interface): 1–11. http://michael.hahsler.net/research/Interface2011/arulesViz/arulesViz.pdf.

Hahsler, Michael, Kurt Hornik, and Thomas Reutterer. 2005. “Implications of Probabilistic Data Modeling for Mining Association Rules.” *Proceedings of the 29th Annual Conference of the Gesellschaft für Klassifikation eV University of Magdeburg March 911 2005* 2(14): 598–605. http://epub.wu-wien.ac.at/dyn/openURL?id=oai:epub.wu-wien.ac.at:epub-wu-01_7f0.

Hastie, Trevor, Robert Tibshirani, and Jerome Friedman. 2001. “Elements of Statistical Learning: Data Mining, Inference and Prediction.” http:// statweb. stanford. edu/~ tibs/ElemStatLearn/.

Hofmann,H, Siebes A, Wilhelm AFX. 2000. “Visualizing Association Rules with Interactive Mosaic.” : 227–35.

Hofmann,Heike,Antony Unwin,and Klaus Bernt. 2001. “The TwoKey Plot for Multiple Association Rules Control.” In *Lecture Notes in Computer Science*, , 472–83.

Hsu, Chung-Chian Hsu Chung-Chian, and Sheng-Hsuan Wang Sheng-Hsuan Wang. 2006. “An Integrated Framework for Visualized and Exploratory Pattern Discovery in Mixed Data.” *IEEE Transactions on Knowledge and Data Engineering* 18(2): 161–73.

Han., J., Cai, Y and Cercone. N, 1993. “Data-Driven Discovery of Quantitative Ruels in Relational Databases.” *IEEE Transactions on Knowledge and Data Engineering* 5(1): 29–40.

Roberto,Jr,Bayardo,J.and Rakesh Agrawa.1999.“Mining the Most Interesting Rules.”: 145–54.

Klemettinen, Mika *et al*. 1994. “Finding Interesting Rules from Large Sets of Discovered Association Rules.” : 401–7.

Lent, B., a. Swami, and J. Widom. 1997. “Clustering Association Rules.” *Proceedings 13th International Conference on Data Engineering*: 220–31.

Klemettinen , M. Toivonen , H. Ronkainen , P. Hätönen, K. Mannila. H.,1995. “Pruning and Grouping Discovered Association Rules.” : 1–6.

Ong, Kian-huat et al. 2002. “CrystalClear: Active Visualization of Association Rules.” *ICDM’02 International Workshop on Active Mining AM2002*.

Rainsford, C P, and J F Roddick. 2000. "Visualisation of Temporal Interval Association Rules." *In Proceedings 2nd International Conference on Intelligent Data Engineering and Automated Learning, (IDEAL*: 91–96.

Valeur, Fredrik. 2006. "Real-Time Intrusion Detection Alert Correlation." *Security* (June): 199.

Yang, Li. 2003. "Visualizing Frequent Itemsets, Association Rules, and Sequential Patterns in Parallel Coordinates." *Computational Science and Its Applications—ICCSA* ...: 21–30. http://www.springerlink.com/index/GU1EX805MNH2GYP2.pdf.