Visual Mining of Association Rules

Dario Bruzzese[1] and Cristina Davino[2]

[1] Dipartimento di Scienze Mediche Preventive, Università di Napoli Federico II
Via S. Pansini, 5 - 80131 Napoli, Italy
dbruzzes@unina.it
[2] Dipartimento di Studi sullo Sviluppo Economico, Università di Macerata
Piazza Oberdan, 3 - 62100 Macerata, Italy
cdavino@unimc.it

Abstract. Association Rules are one of the most widespread data mining tools because they can be easily mined, even from very huge database, and they provide valuable information for many application fields such as marketing, credit scoring, business, etc. The counterpart is that a massive effort is required (due to the large number of rules usually mined) in order to make actionable the retained knowledge. In this framework vizualization tools become essential to have a deep insight into the association structures and interactive features have to be exploited for highlighting the most relevant and meaningful rules.

1 Introduction

A visual data mining approach should complement the data mining techniques because the data visualization allows to understand the process and the models being used. The visualization step is essential when data mining is performed through Association Rules (AR) [2] because of the presence of too many associations where to detect the really relevant implications. It is a matter of fact that even if pruning methods allow to reduce the huge number of mined rules, the resulting subset is often too large for a textual inspection.

Many graphical tools have been proposed in literature (such as [5],[12], [19], [25],[31], [35], [37]) and some of them are implemented in data mining software systems (such as [1], [7], [15], [22], [24], [26], [27], [32], [36]). They make use of classical and basic representations strengthened by interactive features to easily explore the rules. In the following, the main association rules visualizations are discussed trying to highlight their characteristics, their limits and their explanatory features. Most of them represent rules through their characteristics measures (support and confidence) and through the list of involved items and they rarely let to compare items or rules. In this framework two approaches based on Factorial Methods [3] and on Parallel Coordinates [17] are investigated and enhanced.

The analysis of the AR visualizations will be illustrated through a real data set taken from UCI Machine Learning Repository [22].

S.J. Simoff et al. (Eds.): Visual Data Mining, LNCS 4404, pp. 103–122, 2008.

2 Some Issues about Association Rules

AR allow to find frequent patterns and associations in large databases characterized by the presence of a set of transactions, where each transaction is a subset of items. Many field of applications (marketing, credit scoring, business, etc.) require to resort to this data mining tool in order to solve typical problems such as the evaluation of the products assortment, the analysis and the prediction of purchase behaviour of the consumers.

Denoting with $I = i_1, i_2, \ldots, i_m$ a set of m items (e.g. all products bought by a group of customers) and with $T = t_1, t_2, \ldots, t_n$ a set of n transactions (e.g. all products in a customer's basket), an Association Rule R can be expressed in the form $A \rightarrow C$, where both A and C are subset of I such that $A \cap C = \emptyset$. The subset A is the set of antecedent items, also named left hand side (LHS) or body of the rule while C is the set of consequent items, also named right hand side (RHS) or head of the rule (in the following we will denote generic itemsets with capital letters and single items with small letters). The one-to-one rules where both the subset A and the subset C contain only one item $(x \rightarrow y)$ is the simplest association that can be mined but more complex associations can be extracted: many-to-one $(x, ..., y \rightarrow z)$, one-to-many $(x \rightarrow y, ..., z)$ and many-to-many $(x, ..., y \rightarrow z, ..., w)$.

Each rule R is characterized by two measures: the support and the confidence.

The *support* of R can be defined as:

$$S_R = \frac{n_R}{n} \tag{1}$$

where n_R is the number of transactions in T holding $A \cup C$ and it measures the proportion of transactions in T containing both A and C independently from the possible dependence of C from A. In a probabilistic approach the Support is an estimate of the probability of observing in a transaction the items belonging to both the antecedent and the consequence of the rule. The Support can be also referred to a generic itemset if the proportion of transactions sharing the itemset is considered.

The *confidence* of R can be defined as:

$$C_R = \frac{n_R}{n_A} \tag{2}$$

where n_A is the number of transactions in T holding the itemset A. The confidence measures the strength of the implication described by the rule (it is an estimate of the conditional probability of the consequence given the antecedent).

Nowadays, the considerable advances in the computational field allow to analyze many transactions in real time and to easily discover a number of rules that often exceeds the number of transactions. The main drawback of Association Rules is thus the huge number of extracted rules that cannot be manually inspected by the user and the existence of trivial or meaningless associations that are usually mined due to the exhaustive nature of the extraction algorithms. Graphical tools and pruning methods are the main approaches used to face these problems.

Many software tools for the visualization of Association Rules have been proposed in literature. They are limited to visualize only one-to-one rules or many-to-one rules. However the number of displayed rules is so huge that many of them overlap.

AR miners have been sensible to this problem since the introduction of AR in the data mining framework as the abundant literature on pruning methods shows([18], [30], [34]). The first approaches to face the problem of the huge number of discovered AR and of their relevance for the user were based on interestingness measures both subjective and objective. While the former require user domain knowledge and they obviously depend on the user who examines the patterns, the latter force the user to fix a suitable threshold for them. For instance, minimum support and minimum confidence values are usually fixed by the user before mining association rules. Unproper choices of these values may cause many drawbacks: if they are set very low a huge number of rules (some of which being meaningless) will be found. On the contrary, if they are set very high, trivial rules will be found [34]. Moreover, using only confidence and support based thresholds doesn't allow to take into account the strength and the statistical significance of the associations. In this framework, automatic procedure based on statistical tests have been proposed ([6], [14], [21], [23], [33]) even if not all the necessary assumptions are satisfied. The issue is to exploit the theoretical reference framework of the hypothesis tests in order to derive practical criteria, namely score functions [8], to prefer some rules to others. In order to overcome the problems of some of the Association Rules graphical tools, it can be advisable to apply them on the pruned subset of rules.

3 A Real Data Set Application

The data used to discuss the different approaches to Association Rules visualization is taken from UCI Machine Learning Repository [22]. It deals with 101 animals described by 15 boolean attributes (*hair*, *feathers*, *eggs*, *milk*, *airborne*, *aquatic*, *predator*, *toothed*, *backbone*, *breathes*, *venomous*, *fins*, *tail*, *domestic*, *catsize*) and a numeric one (*legs*) which has been categorized in four boolean attributes (0, 2, 4, >4). The zoo data set is a typical machine learning set but it can be useful to highlight the main drawbacks of Association Rules because there is a strong relationship among the animals attributes. Moreover the topic is not a technical one and the rules can be easily understood. The data can be associated to a set of 101 transactions where each transaction is a sub-set of features belonging to an animal.

The number of rules discovered applying the mining process to the set of 101 animals is 3728. This huge number is obtained considering only many-to-one rules at most of the fifth order[1] and fixing a minimum support equal to 0.05

[1] The generated rules contain at most four items in the antecedent and one item in the consequence.

and a minimum confidence equal to 0.5. On the set of mined rules, a sequence of three statistical tests [2] are performed in order to evaluate if each rule significantly satisfies the user specified minimum confidence [21] and support [23] thresholds and if the presence of the itemset in the consequence of each rule depends on the presence on the itemset in the antecedent [6]. At each step, the rules are ranked according to the corresponding test statistics and a subset of rules is obtained pruning the rules out of a suitable threshold. This subset becomes the rules input set for the next step.

The main results of the pruning phase (Table 1) show that after the first step, 18% of the original rules were pruned without influencing significantly the support and confidence ranges. The second step allows to prune a lot of rules (33% of the rules survived at the second step) with not significant support values. Using also the third step as a pruning tool, 1447 final rules remain where the minimum confidence is equal to 0.6 and the minimum support is equal to 0.1. It is worth to notice that the subset of final rules is still very big for a manual inspection and it requires visualization tools to be analysed.

Table 1. Information about the pruning process

	Before Pruning	After Step 1	After Step 2	After Step 3
Nr. of rules	3728	3066	2049	1447
Percentual variation in the nr. of rules		-18%	-33%	-29%
Nr. of involved antecedent items	19	18	18	16
Nr. of involved consequent items	16	16	16	16
Nr. of 2^{nd} order rules	117	83	82	50
Nr. of 3^{rd} order rules	602	471	374	252
Nr. of 4^{th} order rules	1363	1118	758	539
Nr. of 5^{th} order rules	1646	1394	835	606
Minimum Confidence	0.5	0.6	0.6	0.6
Maximum Confidence	1.00	1.00	1.00	1.00
Minimum Support	0.05	0.06	0.1	0.1
Maximum Support	0.73	0.73	0.73	0.73

4 Visualizing Association Rules

Many visualization tools have been introduced in literature and/or implemented in data mining software systems. They differ with respect to the type of represented rules (one-to-one, many-to-one, etc.), to the number of associations that can be visualized, to the type of visualized information (items or measures characterizing the rules), to the number of dimensions (2-D or 3-D) and to the possibility to interact with the graph.

[2] p-values less than 0.01 were considered significant.

4.1 Rule Table

The most immediate Association Rules visualization method is a table (Figure 1) where each row represents a rule and each rule is divided into various parts allocated in different columns of the table. The advantage of this approach is the ability to sort the results by the column of interest. Its main limitation is the close resemblance to the original row textual form so that the user can inspect only few rules without having a global view of all the information.

	A	B	C	D	E	F	G
1	**Antecedent items**				**Consequence**	**Confidence**	**Support**
2	Breathes	Toothed			Backbone	1.00	0.47
3	Backbone	Milk	Toothed		Breathes	1.00	0.40
4	Breathes	Milk	Toothed		Backbone	1.00	0.40
5	0 Legs	Backbone			Tail	0.95	0.18
6	Backbone	Hair	Milk		Breathes	1.00	0.39
7	Breathes	Hair	Milk		Backbone	1.00	0.39
8	Backbone	Breathes	Hair	Toothed	Milk	1.00	0.38
9	0 Legs	Catsize			Tail	0.86	0.06
10	0 Legs	Predator			Eggs	0.76	0.13
11	Eggs	Fins	Predator	Toothed	Tail	1.00	0.09
12	Predator	Tail	Toothed	Venomous	Eggs	0.67	0.02
13	Tail				Toothed	0.69	0.51
14	>4 Legs	Eggs			Breathes	0.67	0.08
15	>4 Legs	Hairborne			Hair	0.67	0.04
16	0 Legs	Aquatic			Backbone	0.94	0.17
17	2 Legs	Aquatic	Eggs		Hairborne	0.83	0.05
18	2 Legs	Aquatic	Tail		Eggs	0.86	0.06

Fig. 1. Rule Table

4.2 Two-Dimensional Matrix

The rules are displayed in a bar diagram where the consequent items are on one axis and the antecedent items on the other axis. The height and the color of the bars are used to represent support and confidence. This visualization approach can be used only in case of one-to-one rules. In figure 2 a subset (50) of the rules extracted on the zoo data set is displayed in a 2-D matrix.

The matrix of associations rules proposed by [14] represent a crushed version of the two dimensional matrix where colors are used to indicate the confidence level while the tone of the colors represents the support.

It is a matter of fact that second order rules are usually pruned (see Table 1) because they represent trivial information. Some softwares like Statistica [29] and Enterprise Miner [27] try without success to overcome this drawback by grouping the items belonging to the antecedent of a rule and by plotting the new unit against the consequence but this strategy is not successful especially when a huge number of rules containing many items in the antecedent is visualized.

4.3 3-D Visualization

The visualization technique proposed by Wong et al. [35] tries to solve the 2-D visualization problems by visualizing many-to-one relationships. The rows of a matrix floor represent the items and the columns represent the rules. Bars with different heights are used to distinguish the consequence and the antecedent of

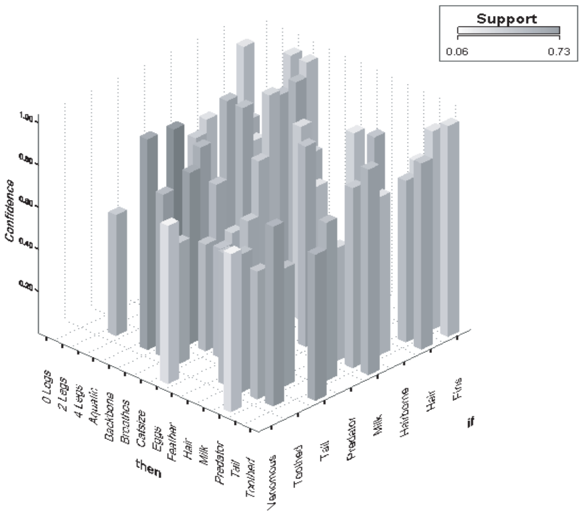


Fig. 2. 2-D matrix representation

each rule. At the far end of the matrix, bars proportional to the confidence and the support measures are represented. The 3-D visualization doesn't impose any limit on the number of items in the antecedent and in the consequence. It allows to analyse the distribution of the association rules and of each item. The 3D view is clear because the support and confidence values are shown at the end of the matrix and in general there is no need for animation.

The visualization proposed by Wong et al. improved 2-D matrix but it still had some problems: the antecedent and consequent items could overlap because they have different positions on the y-axis and the number of displayed rules is limited by the width of matrix floor. In figure 3, 50 rules of different order are plotted using cones instead of bars to partially avoid items overlapping.

4.4 Association Rules Networks

In IBM Intelligent Miner [15] a network representation of AR is provided where each node represents an item and the edges represent the associations. Different colors and width of the arrows are used to represent the confidence and the support. When many rules with many items are represented, the direct graph is not easy to understand because of the superimposition of the edges with the nodes.

Figure 4 shows the visualization of the rules presented in Table 2. If another rule such as $\{$*0 Legs*, *Predator* $\rightarrow$ *Toothed*$\}$ is added to the graph representation, the overlapping among the edges would confuse too much the visualization.

In figure 5 a different network representation is shown [29]. The network displays a subset of 15 rules obtained by setting a maximum order of three, minimum support equal to 50% and minimum confidence equal to 70%. The support

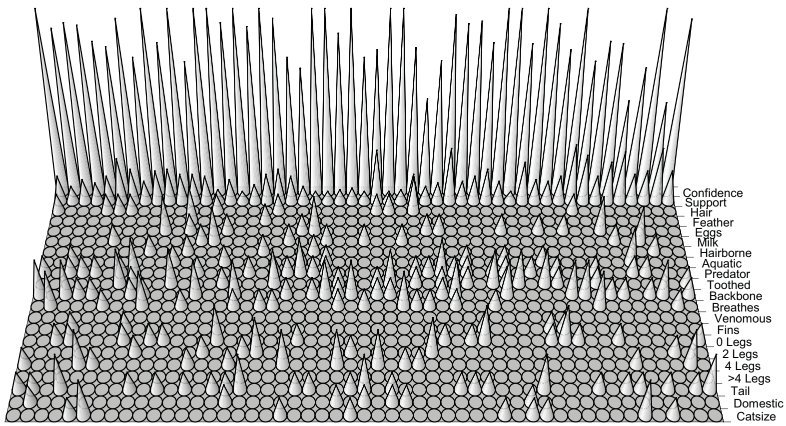


Fig. 3. 3-D matrix representation

Table 2. The rules displayed on the Direct Graph

Antecedent		Consequence	Conf.	Sup.
0 legs		Backbone	0.82	0.18
Aquatic		Backbone	0.80	0.29
0 Legs	Aquatic	Backbone	0.90	0.16
0 Legs		Toothed	0.82	0.18
Backbone		Toothed	0.73	0.6
Backbone		Predator	0.56	0.46

values for the antecedent and consequence of each association rule are indicated by the sizes and colours of each circle. The thickness of each line indicates the confidence value while the sizes and colours of the circles in the center, above the Implies label, indicate the support of each rule. Hence, in figure 5 the strongest support value was found for the one-to-one rules involving the items Tail and backbone. The visualization doesn't allow to identify the most interesting rules especially for the rules with an order greater than 2. The 3D version of the Association Rules Network adds a vertical z - axis to represent the confidence values but as the 2D version, it can be useful only in case of a very small set of rules.

4.5 The TwoKey Plot

The TwoKey plot [31] represents the rules according to their confidence and support values. In such a plot, each rule is a point in a 2-D space where the x-axis and the y-axis ranges respectively from the minimum to the maximum values of the supports and of the confidences and different colors are used to highlight the order of the rules. Many interactive features can facilitate the exploration of the rules such as the selection of a region of the plane where confidence and

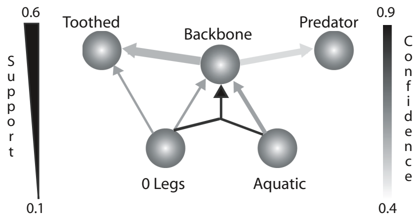


Fig. 4. Direct graph representation

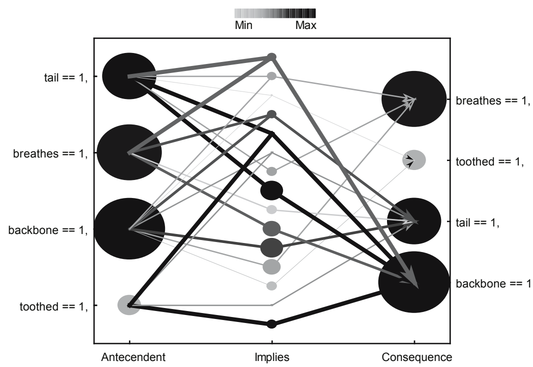


Fig. 5. Association rules network

support are above a user defined threshold or the linking with children, parents and neighbours of a rule. The TwoKey plot can be linked with other displays (barchart of the level, barcharts of the items belonging to sets of rules, mosaic plots [11]).

In Figure 6 a TwoKey plot of one thousand rules extracted from the zoo dataset is shown; an immediate and global overview of the displayed set of rules is provided and it is easy to identify privileged subsets of rules lying in particular regions (for example high confidence rules lining up the top of the graph). The analysis of the items present in the displayed rules necessarily requires to have recourse to the rule table representation which suffers the previously mentioned problems or to a different visualization involving the items.

4.6 Double-Decker Plot

Mosaic plots ([10], [11]) and their variant called Double-Decker ([12], [13], [14]) plots provide a visualization for single association rules but also for all its the related rules. They were introduced to visualize each element of a multivariate contingency table as a tile (or bin) in the plot and they have been adapted to visualize all the attributes involved in a rule by drawing a bar chart for the consequence item and using linking highlighting for the antecedent items. In Figure 7 the double decker plot of the rule *Predator* & *Venomous* & *4 legs* →

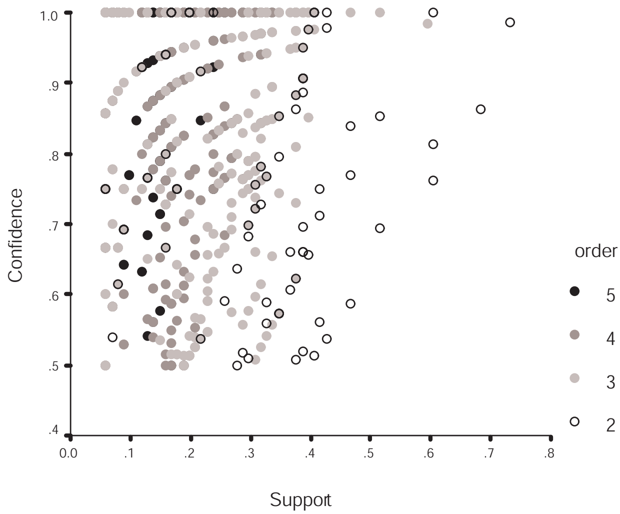


Fig. 6. The TwoKey Plot

Toothed is shown. Each row of the plot corresponds to one item, each gray shade represents one value of this item, the support is the area of highlighting in a bin, the confidence is the proportion of highlighted area in a bin with respect to the total area of the bin. The main drawback of Double Decker plot lies in the possibility to represent one rule at a time or at least all the rules generated from the different combinations of the items belonging to a given rule. In order to have the possibility to represent simultaneously many rules, Hofmann and Wilhelm ([14]) proposed the matrix of Association Rules with and without additional highlighting but only one-to-one rules are taken into consideration.

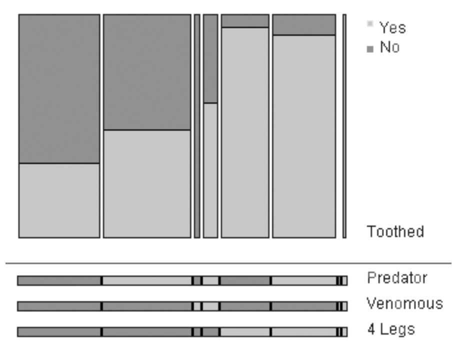


Fig. 7. Double-Decker Plot

4.7 Parallel Coordinates

Parallel coordinates, introduced by Inselberg in 1981 [16], represent a very useful graphical tool to visualize high dimensional data-sets in a two-dimensional space. They appear as a set of vertical axes where each axis describes a dimension of the domain and each case is represented by a line joining its values on the parallel axes.

Parallel coordinates have been used to visualize AR by several authors ([5], [19][3], [37]). The approach proposed by Yang starts from arranging items by groups on a number of parallel axes equal to the maximum order of the rules. A rule is represented as a polyline joining the items in the antecedent followed by an arrow connecting another polyline for the items in the consequence. The items arrangement on each axis should ensure that polylines of itemsets of different groups never intersect with each other. It is a matter of fact that such representation becomes infeasible in case of hundreds or even tens of items and it is not coherent with the original framework of parallel coordinates dealing with quantitative variables.

In figure 8 a parallel coordinate plot of 50 rules of different order is shown. It is evident that in such a case it is not possible to identify disjoint groups of items so that there is an overlapping of the polylines.

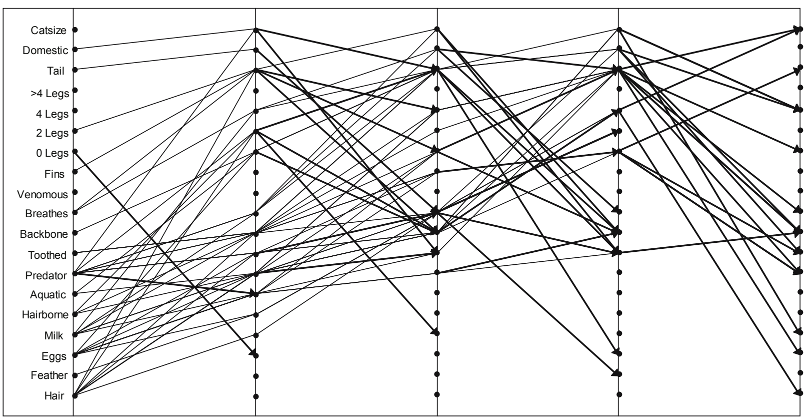


Fig. 8. The parallel coordinate plot proposed by Yang

In the visualization proposed by Bruzzese et al. [5] each antecedent item is a dimension of the graph and it spans according to the utility provided to each rule. The utility of an item i in the antecedent of a rule R is measured by an index called *Item Utility* (IU) based on the comparison between the confidence of the rule with or without item i. Considering the rule $x, y \rightarrow z$ the Item Utility of the y item was defined as follows:

[3] The representation proposed by Kopanakis et al. is not described because it is limited to quantitative AR.

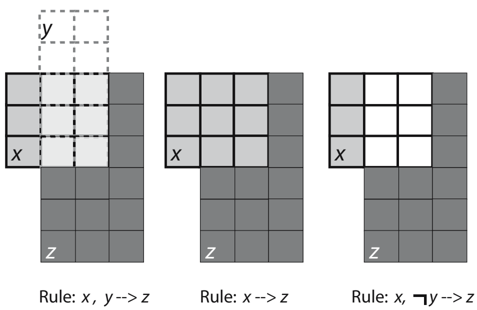
$$IU_y = \frac{C_{x,y\to z} - C_{x\to z}}{max(C_{x\to z}; C_{x,y\to z})} \tag{3}$$

It is a matter of fact that the transactions holding both the x and y items, still contain some transactions sharing the y item. In order to manage the spurious presence of the y item in the rule $x \to z$, it is more appropriate to compare the confidence of the rule $R_1 = x, y \to z$ with the confidence of the rule $R_2 = x, \neg y \to z$ where $\neg y$ denotes the absence of the y item in a transaction. An enhanced *item utility*, called *NIU*, is proposed as follows:

$$NIU_i = \frac{C_R - C_{R(\neg i)}}{max(C_R; C_{R(\neg i)})} \tag{4}$$

where i is a generic antecedent item of the rule R and $R_(\neg i)$ represents the rule R free of the i item.

It results that the *NIU* stresses the importance or the uselessness of an antecedent item with respect to the *IU* as is shown in figure 9 where a graphical comparison among the two indexes is given. Considering each square as a transaction, in figure 9a the confidence of the rule $x, y \to z$ is equal to 1; in 9b the confidence of the rule $x \to z$ is equal to $\frac{2}{3}$ showing that the y item is useful as the confidence decreases when it is not considered. Taking into account the rule $x, \neg y \to z$ (figure 9c) which has a confidence equal to 0, the importance of the y item is highlighted because there are no transactions holding x and z without holding y too.



(a) (b) (c)

Fig. 9. A graphical comparisong beetwen the *IU* and the *NIU* for the y item

The *NIU* ranges in the interval $]-1; 1]$. The value -1 is not included in the interval as it refers to rules with confidence equal to 0 which can never be mined. If $NIU_i \in\,]-1; 0[$, the i item is harmful and the rule can be pruned as the intersection between the i item and the other antecedent items is not relevant for the prediction of the consequence. If $NIU_i \in\,]0; 1]$, the item is useful as its interaction with the other antecedent items improves the capability to explain the consequence. The case NIU=0 refers to the presence of a redundant item as its presence in a transaction doesn't add further information. From a

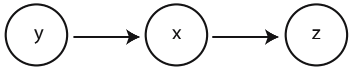


Fig. 10. The graph representation of the rule $x, y \rightarrow z$ when z and y are conditionally independent given x

probabilistic point of view the case $NIU = 0$ for the y item in the rule $x, y \rightarrow z$ means that z and y are conditionally independent given x and the rule can be represented as a directed acyclic graph (Figure 10) where the link between the item y and the item z goes through x (metaphorically x screens off z form y).

A view of the discovered rules can be obtained plotting on parallel coordinates the *NIU* of each item belonging to the antecedent of a rule.

Some of the interaction tools of parallel coordinates [17] are exploited in order to visualize, interpret and reduce the number of rules. In particular, data analysis can be facilitated by:

- selecting a subgroup of rules with one or more items below a specified *NIU* threshold in order to remove selected lines from the plot;
- identifying axes (items) with very dense positive values, given a consequence, in order to highlight items with a high explicative power;
- adding two supplementary dimensions corresponding to the support and confidence of the rules in order to remove those rules with values of these parameters below a specified threshold;
- selecting high confidence rules in order to identify sets of items involved in very strong associations;
- changing the order of the dimensions on the basis of *NIU* distributions.

In figure 11 a plot of 736 rules with a common consequence (*Toothed*) is shown. Each rule is represented as a line joining the axis corresponding to its antecedent items, to its confidence and support values. The most explicative items (*0 Legs*, *4 Legs*, *Backbone*, *Fins*, *Milk*) and the most critical items (*2 Legs*, *Eggs*) can be easily identified respectively as the ones with very dense positive or negative *NIU* values.

The empirical evaluation of the item utility must be accomplished with the assessment of its statistical significance in order to obtain an overall measure of the importance of each item in a rule. At this aim a statistical test is introduced to verify whether the difference between the two confidences is equal or greater than 0. Let C_{R_1} be the confidence of the rule $R_1 = x, y \rightarrow z$ and C_{R_2} be the confidence of the rule $R_2 = x, \neg y \rightarrow z$. The test is performed starting from the following hypothesis:

$$H_0 : C_{R_1} = C_{R_2} \qquad H_0 : C_{R_1} > C_{R_2} \tag{5}$$

Under the null hypothesis the test statistics T_{NIU}:

$$T_{NIU} = \frac{C_{R_1} - C_{R_2}}{\sqrt{C^*(1 - C^*)\left(\frac{1}{n_{x,y}} + \frac{1}{n_{x,\neg y}}\right)}} \tag{6}$$

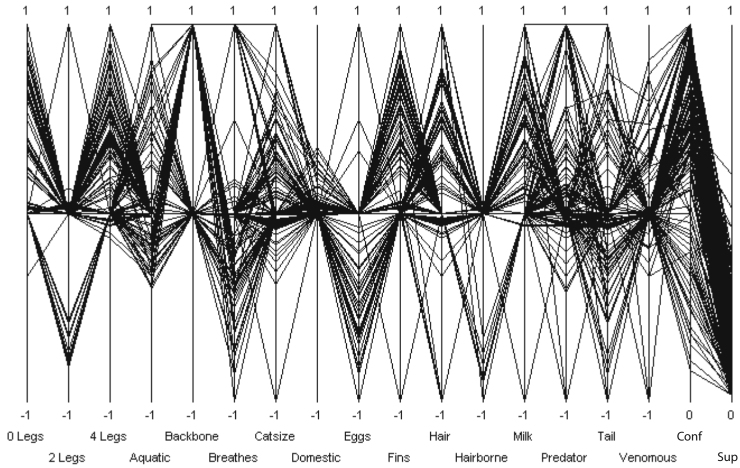


Fig. 11. The parallel coordinates plot of rules with consequence equal to *Toothed*

approximates a standard normal distribution given that $n_{x,y}$ and $n_{x,\neg y}$ are sufficiently large. The term C^* refers to the estimate of the conjoint proportion:

$$C^* = \frac{n_{x,y,z} + n_{x,\neg y,z}}{n_{x,y} + n_{x,\neg y}} \tag{7}$$

From equation 7 it follows that $C*$ measures the confidence of the rule $R^* = x \rightarrow z$.

When we deal with one-to-one rules, the test statistics T_{NIU} given in equation 6 is equal to the *Difference of Confidence (Doc)* test statistic proposed in [14] where the confidence of a rule is compared with the confidence of the rule obtained considering the same consequence but the negation of the whole antecedent set of items. The test can be used to prune those rules where at least one antecedent item has a *NIU* not significantly greater than 0 because the interaction among all the antecedent items is not relevant and a lower order rule must be retained.

Figure 12 shows a parallel plot of the 60 rules that survived the test with a significance level of 0.05. The set of rules is characterised by high confidence values and by a strong interaction among the shared items, with *NIU* values often equal to 1.

4.8 Factorial Planes

As a matter of fact, the number of extracted rules, and even the number of rules after pruning, are huge, which makes manual inspection difficult. A factorial method can be used to face this problem because it allows to synthesize the information stored in the rules and to visualize the associations structure on 2-dimensional graphs.

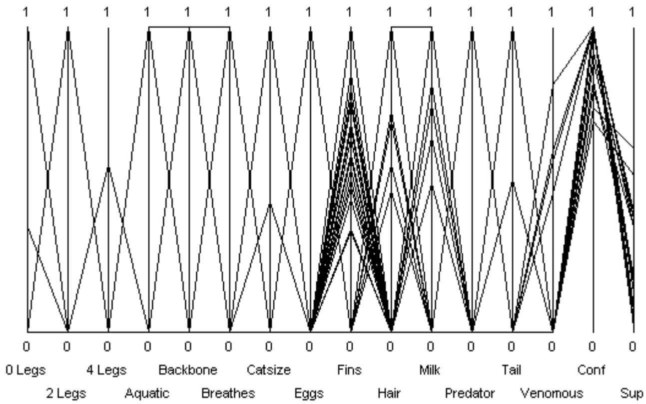


Fig. 12. The parallel coordinate plot of rules with consequence equal to *Toothed* after pruning

The rules being synthesized are stored in a data matrix where the number of n rows is equal to the number of rules and the number of columns ($p = p_{if} + p_{then}$) corresponds to the total number of different items, both in the antecedent part (p_{if}) and in the consequent part (p_{then}) of the n rules. Each rule is coded by a binary array assuming value 1 if the corresponding column item is present in the rule and value 0 otherwise. The well known confidence and support measures are also considered as added columns. The final data matrix has thus $n \times (p_{if} + p_{then} + 2)$ dimensions and it can be analysed through the Multiple Correspondence Analysis (MCA) ([3], [9]) that allows to represent the relationships among the observed variables, the similarities and differences among the rules and the interactions between them. MCA allows to reduce the number of original variables finding linear combinations of them, the so called *factors*, that minimize the deriving loss of information due to the dimensionality reduction. Different roles are assigned to the columns of the data matrix: the antecedent items are called *active* variables and they intervene directly in the analysis defining the factors; the consequent items and the support and the confidence values are called *supplementary* variables because they depend from the former and are projected later on the defined factorial planes.

Referring at the zoo data set, the rules survived to a pruning process [6] are 1447 and they involve 16 different items[4] both in the antecedent part (p_{if}) and in the consequence (p_{then}). The set of rules should thus be represented in a 16-dimensional space and the set of items in a 1447-dimensional space. In order to reduce the number of original variables through the factors, it is necessary to evaluate the loss of information deriving or the variability explained by the retained factors. According to the Benzcri approach [3] for the evaluation of the explained variability in case of MCA, in table 3, the explained variability and the cumulative variability is shown. The first two factors share more than the

[4] *Venomous*, *Domestic* and *>4 Legs* are the items removed by the pruning procedure.

Table 3. Total inertia decomposition

Factor	% of variability	Cumulative %	
1	44	44	**
2	40	84	************************************
3	12	96	***********
4	4	100	****

80% of the total inertia and they correspond to the highest change in level in the percentage of variability.

Once the MCA is performed it is possible to represent the rules and the items on reduced dimensions subspaces: the factorial planes allowing to explain at least a user defined threshold of the total variability (in the zoo example, the first factorial plane) or a user defined factorial plane or the factorial plane best defined by a user chosen item.

Different *views* on the set of rules can be obtained exploiting the results of the MCA.

1. **Items Visualization.** A graphical representation of the antecedent and the consequent items is provided by the factorial plane where the item points have a dimension proportional to their supports and the confidence and the support are represented by oriented segments linking the origin of the axes to their projection on the plane. In Figure 13 the active and the supplementary items are visualized together with the confidence and support arrows.

   Privileged regions characterized by strong rules can be identified in case of high coordinates of the confidence and the support because their coordinates represent the correlation coefficients with the axes.

   The proximity between two antecedent items shows the presence of a set of rules sharing them while the proximity between two consequent items is related to a common causal structure. Finally, the closeness between antecedent items and consequent items highlights the presence of a set of rules with a common dependence structure.
2. **Rules Visualization.** Another view on the mined knowledge is provided by the rules representation on the factorial plane. Graphical tools and interactive features can help in the interpretation of the graph: the rules are represented by points with a dimension proportional to their confidence, the proximity among two or more rules shows the presence of a common structure of antecedent items associated to different consequences, a selected subset of rules can be inspected in a tabular format. For example in table 4 the subset of the rules selected in figure 14 is listed.

   It is worth of notice that this subset of rules is very close on the plane because they have similar antecedent structures sharing at least one item, even some rules overlap because they have exactly the same antecedent structure.

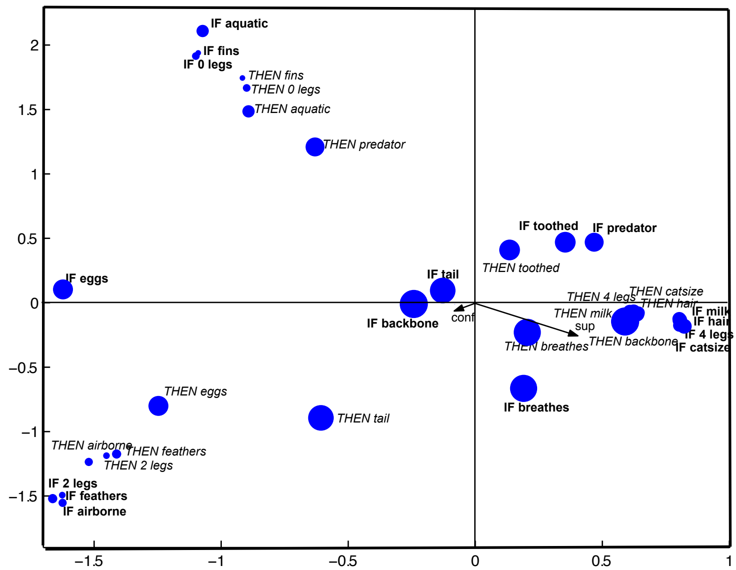


Fig. 13. The items representation

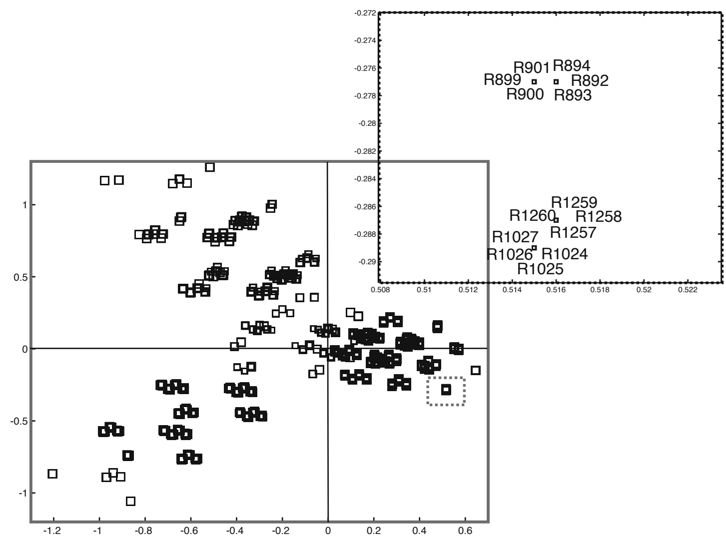


Fig. 14. The rules representation

It is possible to imagine to transform the set of overlapping rules into a higher order macro-rule obtained linking the common behaviour described by the antecedent items to the logical disjunction of the different consequent items.

Table 4. Description of a subset of overlapping rules

Rule	Antecedent	Consequence	Conf.	Sup.
899	Hair & Milk & Breathes & Catsize	Toothed	0.97	0.29
900	Hair & Milk & Breathes & Catsize	Backbone	1.00	0.30
901	Hair & Milk & Breathes & Catsize	4 legs	0.83	0.25
892	Hair & Milk & Breathes & 4 legs	Toothed	0.97	0.30
893	Hair & Milk & Breathes & 4 legs	Backbone	1.00	0.31
894	Hair & Milk & Breathes & 4 legs	Tail	0.90	0.28
1257	Milk & Breathes & 4 legs & Catsize	Hair	1.00	0.25
1258	Milk & Breathes & 4 legs & Catsize	Toothed	0.96	0.24
1259	Milk & Breathes & 4 legs & Catsize	Backbone	1.00	0.25
1260	Milk & Breathes & 4 legs & Catsize	Tail	0.92	0.23
1024	Hair & Breathes & 4 legs & Catsize	Milk	1.00	0.25
1025	Hair & Breathes & 4 legs & Catsize	Toothed	0.96	0.24
1026	Hair & Breathes & 4 legs & Catsize	Backbone	1.00	0.25
1027	Hair & Breathes & 4 legs & Catsize	Tail	0.92	0.23

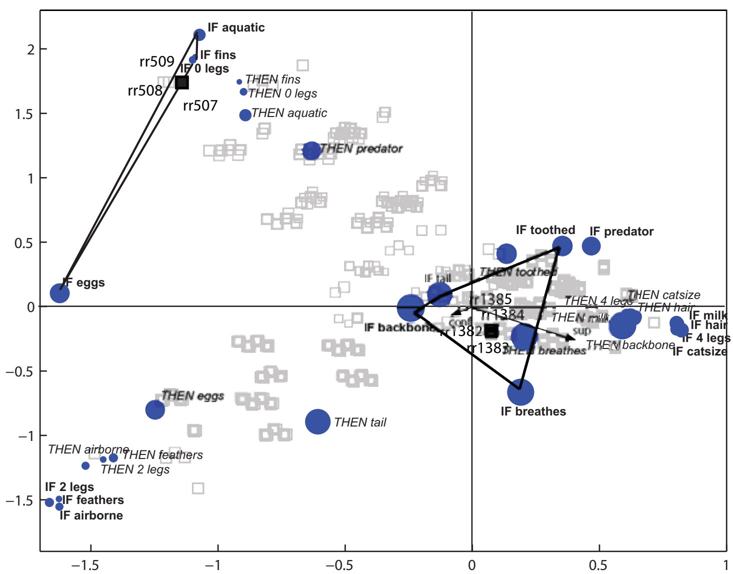


Fig. 15. The Conjoint representation

3. **Conjoint Visualization.** The factorial planes features allow to visualize simultaneously the items and the rules. In the conjoint representation, aside from a scale factor, each rule is surrounded by the antecedent items it holds and vice versa each item is surrounded by the rules sharing it. By linking two or more active items it is possible to highlight all the rules that contain at least one of the selected items in the antecent. For example in figure 15

two groups of rules have been closed inside the polygons joining the items they share in the antecedent.

5 Concluding Remarks

Association Rules Visualization is emerging as a crucial step in a data mining process in order to profitably use the extracted knowledge. In this paper the main approaches used to face this problem have been discussed. It rises that, up to day, a compromise have to be done between the quantity of information (in terms of number of rules) that could be visualized and the depth of insight that can be reached. This suggests that there is not a winning visualization but their strength lies in the possibility to exploit the synergic power deriving from their conjoint use. Moreover, it is advisable a stronger interaction among the visualization tools and the data mining process that should incorporate each other.

Acknowledgements. The paper was financially supported by University of Macerata grant (2004) *Metodi Statistici Multivariati per l'Analisi e la Valutazione delle Performance in Campo Economico e Aziendale.*

References

1. Advanced Visual Systems (AVS), OpenViz. `http://www.avs.com/software/`
2. Agrawal, R., Imielinski, T., Swami, A.: Mining Association Rules between Sets of Items in Large Databases. In: Proceedings of the 1993 ACM SIGMOD Conference, Washington DC, USA, pp. 207–216 (May 1993)
3. Benzècri, J.-P.: L'Analyse des Donnèes, Dunod, Paris (1973)
4. Bruzzese, D., Buono, P.: Combining Visual Techniques for Association Rules Exploration. In: Proceedings of the International Conference Advances Visual Interfaces, Gallipoli, Italy, May 25-28 (2004)
5. Bruzzese, D., Davino, C., Vistocco, D.: Parallel Coordinates for Interactive Exploration of Association Rules. In: Proceedings of the 10th International Conference on Human - Computer Interaction, Creta, Greece, June 22-27. Lawrence Erlbaum, Mahwah (2003)
6. Bruzzese, D., Davino, C.: Significant Knowledge Extraction from Association Rules. In: Electronic Proceedings of the International Conference Knowledge Extraction and Modeling Workshop, Anacapri, Italy, September 4-6 (2006)
7. Clementine, Suite from SPSS, `http://www.spss.com/Clementine/`
8. Glymour, C., Madigan, D., Pregibon, D., Smyth, P.: Statistical Inference and Data Mining. Communications of the ACM (1996)
9. Greenacre, M.: Correspondence Analysis in Practice. Academic Press, London (1993)
10. Hartigan, J., Kleiner, B.: Mosaics for contingency tables. In: Proceedings of the 13th Symposium on the interface, pp. 268–273 (1981)
11. Hofmann, H.: Exploring categorical data: interactive mosaic plots. Metrika 51(1), 11–26 (2000)

12. Hofmann, H., Siebes, A., Wilhelm, A.: Visualizing Association Rules with Interactive Mosaic Plots. In: Proceedings of the 6th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 227–235 (2000)
13. Hofmann, H., Wilhelm, A.: Validation of Association Rules by Interactive Mosaic Plots. In: Bethlehem, J.G., van der Heijden, P.G.M. (eds.) Compstat 2000 - Proceedings in Computational Statistics, pp. 499–504. Physica-Verlag, Heidelberg (2000)
14. Hofmann, H., Wilhelm, A.: Visual Comparison of Association Rules. Computational Statistics 16, 399–416 (2001)
15. IBM Intelligent Miner for Data, `http://www.software.ibm.com/data/intelli-mine`
16. Inselberg, A.: N-dimensional Graphics, part I - Lines and Hyperplanes, in IBM LASC Tech. Rep. G320-2711, 140 pages. IBM LA Scientific Center (1981)
17. Inselberg, A.: Visual Data Mining with Parallel Coordinates. Computational Statistics 13(1), 47–64 (1998)
18. Klemettinen, M., Mannila, H., Ronkainen, P., Toivonen, H., Verkamo, A.I.: Finding interesting rules from large sets of discovered association rules. In: Proceedings of the Third International Conference on Information and Knowledge Management CIKM 1994, pp. 401–407 (1994)
19. Kopanakis, I., Theodoulidis, B.: Visual Data Mining & Modeling Techniques. In: 4th International Conference on Knowledge Discovery and Data Mining (2001)
20. Liu, B., Hsu, W., Ma, Y.: Pruning and Summarizing the Discovered Associations. In: Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD 1999), San Diego, CA, USA, August 15-18 (1999)
21. Liu, B., Ma, Y., Lee, R.: Analyzing the Interestingness of Association Rules from Temporal Dimensions. In: International Conference on Data Mining, CA (2001)
22. Machine learning, `http://www1.ics.uci.edu/~mlearn/MLSummary.html`
23. Megiddo, N., Srikant, R.: Discovering Predictive Association Rules. In: Knowledge Discovery and Data Mining (KDD 1998), pp. 274–278 (1998)
24. Miner3D, Miner3D Excel, `http://www.miner3d.com/m3Dxl/`
25. Ong, K.-H., Ong, K.-L., Ng, W.-K., Lim, E.-P.: CrystalClear: active Visualization of Association Rules. In: Proc. of the Int. workshop on Active Mining, Japan (2002)
26. Purple Insight Mineset, `http://www.purpleinsight.com`
27. Sas Enterprise Miner, `http://www.sas.com/technologies/analytics/datamining/miner`
28. Shah, D., Lakshmanan, L.V.S., Ramamritham, K., Sudarshan, S.: Interestingness and Pruning of Mined Patterns. In: Workshop Notes of the 1999 ACM SIGMOD Research Issues in Data Mining and Knowledge Discovery (1999)
29. Statistica, `http://www.statsoft.com`
30. Toivonen, H., Klemettinen, M., Ronkainen, P., Hatonen, K., Mannila, H.: Pruning and grouping of discovered association rules. In: Workshop Notes of the ECML-95 Workshop on Statistics, Machine Learning, and Knowledge Discovery in Databases, Heraklion, Greece, April 1995, pp. 47–52 (1995)
31. Unwin, A., Hofmann, H., Bernt, K.: The TwoKey Plot for Multiple Association Rules Control. In: Siebes, A., De Raedt, L. (eds.) PKDD 2001. LNCS (LNAI), vol. 2168. Springer, Heidelberg (2001)
32. VisualMine, `http://www.visualmine.com/`
33. Webb, G.I.: Preliminary Investigations into Statistically Valid Exploratory Rule Discovery. In: Proceedings of the Australasian Data Mining Workshop, Sydney (2003)

34. Weber, I.: On Pruning Strategies for Discovery of Generalized and Quantitative Association Rules. In: Proceedings of Knowledge Discovery and Data Mining Workshop, Singapore (1998)
35. Wong, P.C., Whitney, P., Thomas, J.: Visualizing Association Rules for Text Mining. In: Wills, G., Keim, D. (eds.) Proceedings of IEEE Information Visualization 1999. IEEE CS Press, Los Alamitos (1999)
36. XGvis: A System for Multidimensional Scaling and Graph Layout in any Dimension, http://www.research.att.com/areas/stat/xgobi/
37. Yang, L.: Visualizing Frequent Itemsets, Association Rules and Sequential Patterns in Parallel Coordinates. In: Kumar, V., Gavrilova, M.L., Tan, C.J.K., L'Ecuyer, P. (eds.) ICCSA 2003. LNCS, vol. 2667, pp. 21–30. Springer, Heidelberg (2003)