Computational Statistics (2001) 16:399–415

Computational Statistics

Visual Comparison of Association Rules

Heike Hofmann and Adalbert Wilhelm

Institute of Mathematics, Augsburg University, Universitätsstr. 14, 86135 Augsburg, Germany

Summary

Rule induction methods are widely applied tools for mining large data bases. They are often used as a starting point in undirected data mining, i.e. when you do not know what specific patterns to look for. One form of rule induction methods are association rules which have their origin in market basket analysis. Since an evaluation of their results is often hard due to the mass of output, pruning methods are needed to turn the output of association rules into a manageable number of patterns. We will present some statistical measures and their depictions that are useful to assess the quality of an association rule. We will show plots portraying confidence and support for individual rules as well as for sets of rules as alternatives to displays currently used in commercial software. A new quality measure for association rules, the *doc*, will be introduced which overcomes some of the problems of support and confidence of association rules and can be visualised even for hundreds of rules in one diagram.

Keywords: Association Rules, confidence, Double-decker plots, support, Visualization of Association Rules

1 Motivation

Association rules (Agrawal et al. 1993) have developed to a well established tool in datamining software for finding associations among explanatory variables. One of the problems of this technique is, however, that a vast amount of rules is generated in the mining process - discovering "true" results among them remains still a hard task.

Association rules are well-known in market basket analysis and aim in identiying items that occur together more frequently than might be expected if the items were independent of one-another. In data base terminology an association rule is written in the form :

$$A \rightarrow B \ (supp\ , conf\,)$$

in which A and B are selections with disjoint attribute sets and *supp* and *conf* denote the *support* and the *confidence* of the rule. In statistical terms association rules can be described with the help of binary random variables $X_j : \Omega \rightarrow \{0,1\}$ $(1 \leq j \leq p)$ that are defined on some sample space Ω, a subset of the set of positive integers, the power set $\mathcal{P}(\Omega)$ and some probability measure P on the measurable space $(\Omega, \mathcal{P}(\Omega))$. In the context of association rules we are mainly interested in the occurence of events, that is, we are looking for those instances $\omega \in \Omega$ in which some variable X_j takes on the value '1'. For convenience we use the same notation X_j for both the random variable as well as the event $\{X_j = 1\}$. The complementary event $\{X_j = 1\}$ is denoted by $\neg X_j$ and we set $P(X_j) = P(\{omega \in \Omega : X_j(\omega) = 1\})$. Univariate events are of minor interest, our focus is directed towards occurence of multivariate events $X_{j_1} = 1, X_{j_2} = 1, \ldots, X_{j_\ell} = 1$ which we abbreviate with the notation $\{X = 1\}$ or simply X. Thus, $P(X) = P(\{\omega \in \Omega : X_{j_1} = 1, \ldots, X_{j_\ell} = 1\}$. Given two events X and Y the support *supp* $(X \rightarrow Y)$ of an association rule corresponds to the joint probability with which the two events X and Y occur and the *conditional propability* of event Y given event X corresponds to the confidence *conf* $(X \rightarrow Y)$.

The typical procedure for "mining" associations among variables is then to generate **all** possible rules with minimum support and minimum confidence, which means

$$X \rightarrow Y \iff \begin{cases} P(X \cap Y) \geq minsupp \\ P(Y|X) \geq minconf. \end{cases}$$

This is an exponentially explosive undertaking but a variety of algorithms has been proposed to achieve this goal, many of them being modifications of the a-priori algorithm by Agrawal et al. (1993). To filter between "interesting" and less interesting results, several measures of "interestingness" have been introduced in the literature (Srikant & Agrawal (1995), Bayardo & Agrawal (1999)). They find their statistical counterpart in several variations of gaussian and χ^2-tests.

The main point is to elicit those rules which have been previously unknown, which show an interesting and unexpected structure, and which can be exploited. The final goal of association rules is typically a segmentation of the population under consideration. Therefore, pruning methods should aim to result in a precinct description of the relevant subpopulations. There are several approaches for pruning association rules. Among them pruning based on ancestor/successor relationships among the rules (Srikant & Agrawal 1995), other authors propose the use of templates (Klemettinen et al 1994) or calculate covers of rules (Toivonen et al 1995). One aspect, which "blind" pruning methods do not regard, though, is the explicability of results - interpreting the association rules found is of great interest. However, while one association rule may help the analyst to link the result with some kind of meta-information, which explains the "why" of an association rule, it may be inferior to another association rule, which describes the same situation, in the eyes of a pruning method. This situation occurs particularly often when the variables are highly correlated.

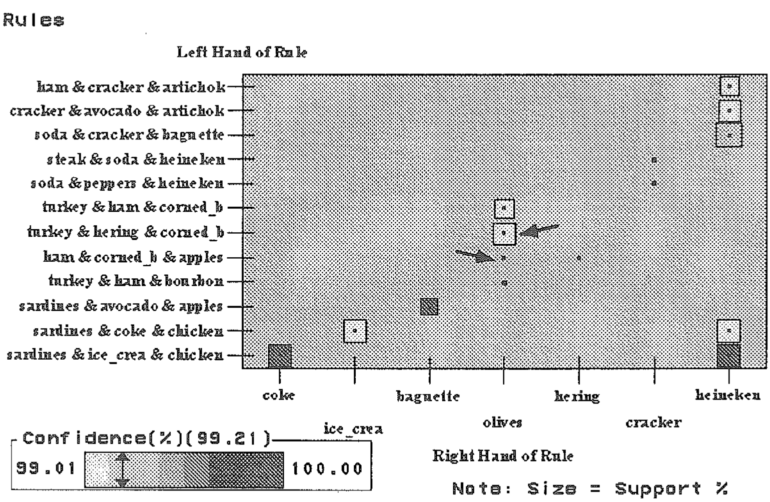


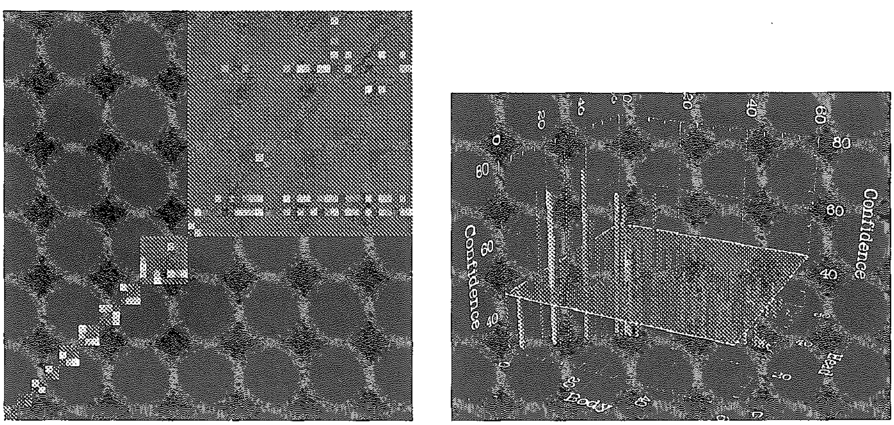
Figure 1: *SAS Enterprise Miner: visualization of all 15 association rules with a minimum confidence of 99%.*

In this paper we propose several graphical methods for getting an overview of a set of association rules without loosing the information necessary for explaining results. We will concentrate in the sequel on confidence and support as measures of quality of association rules, although we are aware of the drawbacks (see Hofmann, Siebes & Wilhelm (2000)) of these measurements. We illustrate our procedures by the use of a market basket analysis example data file that comes with the SAS Enterprise Miner (in the following referred to as the *SAS Assocs Data*). In this data set the shopping behaviour of 2001 customers with respect to 20 items has been registered.

There have been only a few attempts to use visual tools for displaying the quality of rules and to use these visual displays for pruning. Figure 1 shows a rather typical graphic for the visualization of association rules in commercial software packages. This example is taken from the SAS Enterprise Miner. It shows 15 rules in a matrix, where each row corresponds to one left-hand-side of a rule and each column corresponds to one right-hand-side. Each rule, which fulfils *minsupp* and *minconf* is drawn as a square. The different grey shades of the tiles are assigned by the confidences of the rules - in Figure 1 the confidences vary between 99% (light grey) and 100 % (dark grey). The size of the squares is given by the support of the corresponding rule. In fact, the area is proportional to the square of the support - which visually is problematic, since for instance the following two rules (marked by the arrows in Figure 1)

$$turkey \ \& \ herring \ \& \ corned \ beef \rightarrow olives$$
$$ham \ \& \ corned \ beef \ \& \ apples \rightarrow olives$$

have support 11.19% and 3.1%, respectively. So, the measures differ by a factor of approximately 4, whereas the areas differ by a factor of 16. This difference between effect within the data and effect within the visual display has been mentioned a lot in the literature of visualization techniques and has been named the *lie-factor* by Tufte (1983).



a) 2-D Visualization b) 3-D Visualization

Figure 2: *2 and 3 D Visualization of association rules with two items.*

Another commercial datamining software, the Intelligent Miner by IBM, provides several possibilities to visualise association rules. Figure 2 shows two of them for the visualization of 2-item rules (i.e. one item in the body of the rule, one in the head of it). Plot a) is similar to the approach of the SAS Enterprise Miner (cf. Figure 1). It shows a matrix of association rules with two items each, the colour denotes different confidence levels. Plot b) shows the same rules in a three-dimensional display. Each association rule is drawn as a bar, starting from a matrix of all combinations of body and head items. The height of each bar denotes the confidence, the colour gives the support of the rule. Both approaches have in common, that they use one dimension in addition to the dimensionality of the objects: for visualising 2-D objects three dimensions are used, for 3-D objects four dimensions are used (we have to count colour as an additional dimension). This, however, is not necessary. We want to show how the available dimensions can be used more efficiently.

2 Plotting confidence vs. support

An obvious way of displaying both criteria, confidence and support, in one plot is to display them in a scatterplot. Although this seems rather simple it looks as if nobody else has ever done this, otherwise the particular structure visible in such plots would have been noticed earlier. In such scatterplots all points appear on rays exiting from the origin (see Figure 3). The reason for

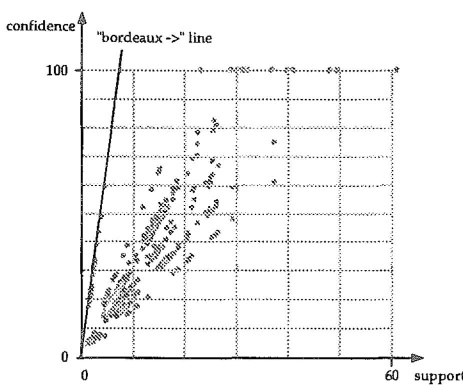


Figure 3: *Confidence vs. support of all 2 item rules in the Assocs Data.*

this is simple. Let $X \to Y$ be an association rule, then

$$\frac{conf\,(X \to Y)}{supp\,(X \to Y)} = \frac{P(Y \mid X)}{P(X \cap Y)} = \frac{1}{P(X)},$$

i.e. for each left hand side X, there is a straight line, on which all data points $X \to Y$ appear with increasing confidences (starting from the origin). The points at the top of the display with 100% confidence indicate mainly rules of the form $X \to X$ - the single items therefore are sorted according to their sizes (smallest is bordeaux on the very left, largest single item is heineken - on the very right side). Direct comparisons between rules, which lie on vertical or horizontal lines are now possible: if two rules have the same confidence,

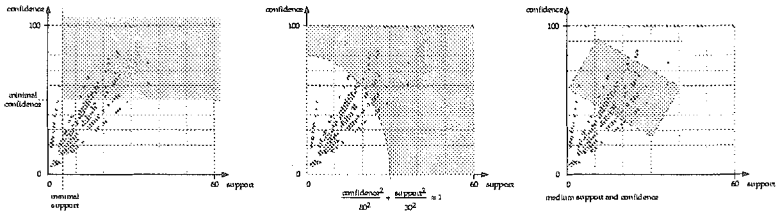


Figure 4: *Two different principles of accepting rules. On the left, the traditional area of acceptance with specified minimal confidence and minimal support. On the right a possible alternative.*

the one with the higher support is better (the same is true for two rules with the same support). Accepted rules therefore (see sketch in Figure 4, left side) fall in a rectangle bounded by *minconf* and *minsupp*. Alternatively, one can also think of a different criterion for accepting rules. The plot in the middle of Figure 4 shows an area of acceptance, where a low support can be balanced by a high confidence and vice versa. Using interactive selection methods and linking opens a lot more choices for criteria of acceptance, all of which may be sensible in the background of a specific application. From a practitioner's viewpoint rules with high confidence and support are not very interesting because they are already known to the domain experts and hence it is more desirable to focus on medium confidence and support levels, as is indicated by the acceptance area in the plot at the right hand side of Figure 4.

3 Doubledecker Plots

Hofmann, Siebes & Wilhelm (2000) propose to use mosaic plots to visualize all possible combinations of explanatory variables involved in a rule. By drawing a bar chart for the response Y, selecting the '1'-category for Y, and using linked highlighting every bin in the mosaic plot shows a different association rule of the form

$$X_1 = x_1 \wedge X_2 = x_2 \wedge \ldots \wedge X_k = x_k \rightarrow Y = 1,$$

where any of the x_i is either 0 or 1. The support of an association rule is represented by the area of highlighting in the corresponding bin, the confidence of the rule can be deduced from the proportion of the highlighted area in a bin and the total area of the bin. Using a different layout for the bins, mosaic plots can be turned into doubledecker plots which make labelling possible and in which the confidence of a rule is given by the height of the highlighted area in a bin, see Hofmann & Wilhelm (2000) for details. In Figure 5 and Figure 6 two examples for doubledecker plots are given that correspond to a strong and a weak association rule, respectively. Note that both rules have approximately the same confidence and the same support as given in the following table:

rule	conf.	supp.
heineken & coke & chicken → sardines	98.31	11.59
cracker & soda & olives → heineken	96.83	12.18

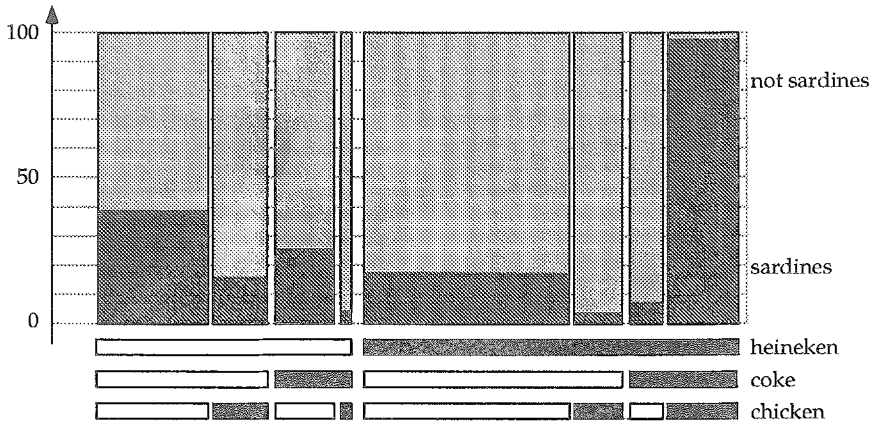


Figure 5: *A doubledecker plot showing a strong association between* `heineken` *&* `coke` *&* `chicken` *on one side and* `sardines` *on the other hand.*

Figure 5 shows the rule `heineken & coke & chicken` → `sardines` which is strong because the corresponding bin (the right-most bin in the doubledecker plot) is the only one that is filled to a high extent by the highlighting colour.

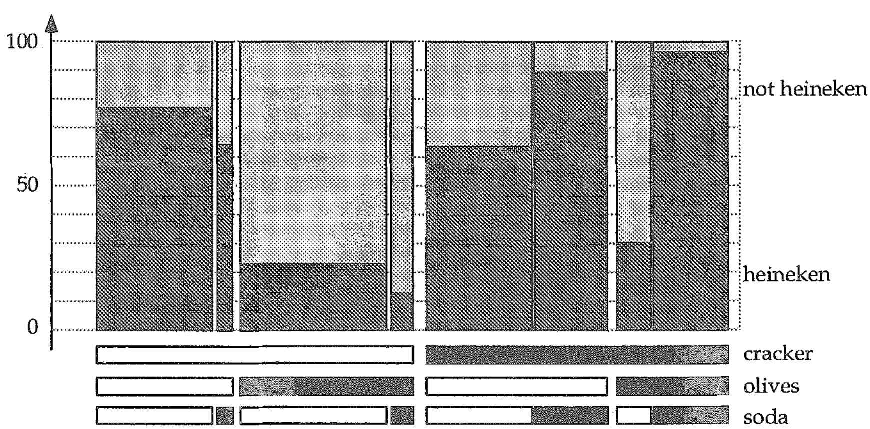


Figure 6: *A doubledecker plot showing a weak association rule* `cracker` *&* `soda` *&* `olives` → `heineken`*.*

In Figure 6 in contrast several bins with large highlighting heights can be identified and therefore the rule **cracker & soda & olives → heineken** which is represented by the right-most bin is not specific for the underlying subpopulation. In contrast, there are quite a few combinations of the variables involved that define the same subgroup.

4 Difference of confidences - DOC

Association rules tend to prefer frequent events as their response. This implies that for any large event Y that exceeds the minimum confidence an association rule $X \to Y$ will be generated as long as the intersection $X \cap Y$ exceeds the minimum support. Such a rule $X \to Y$, which is based only upon the fact that $P(Y)$ is large is not reliable at all since X and Y can be close to being statistically independent.

The *(model-) lift* (Piatetsky-Shapiro& Masand 1999) of a rule $X \to Y$ is defined as

$$lift(X \to Y) := \frac{conf\,(X \to Y)}{supp\,(Y)}.$$

and compares a model with the random situation. The lift, thus, is a measure of the deviation from independence of the two events X and Y, since it can be written as

$$lift(X \to Y) = \frac{P(Y \mid X)}{P(Y)} = \frac{P(X \cap Y)}{P(X) \cdot P(Y)}.$$

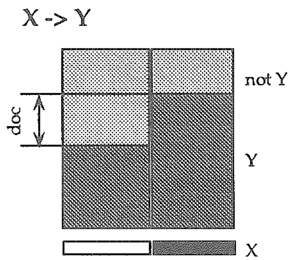
Instead of comparing an association rule with the random situation we can measure how much the variable X contributes to the prediction of Y. One way to do this is to measure the differences of confidences *doc* for the rule $X \to Y$ and $\neg X \to Y$. We define:

$$doc(X \to Y) := conf\,(X \to Y) - conf\,(\neg X \to Y).$$

The $doc(X \to Y)$ can be re- expressed as:

$$doc(X \to Y) = \frac{P(X \cap Y) - P(X) \cdot P(Y)}{P(X) \cdot P(\neg X)}.$$

It can be shown that the product $doc(X \to Y) \cdot doc(Y \to X)$ is approximately χ^2- distributed. The doc measure can be seen directly in a doubledecker plot, as the following sketch shows:



The support of $X \rightarrow Y$ is directly proportional to the highlighted rectangle on the right side, its height gives the confidence of $X \rightarrow Y$.
Similarly, the height of the highlighted rectangle on the left side shows the confidence of rule $\neg X \rightarrow Y$.
The difference of these confidences gives the *doc*.

Figure 7 shows an example of the following six association rules:

Rule			support	confidence	lift	doc
heineken	$\rightarrow$	corned beef	0.12	0.21	0.54	-0.45
apples	$\rightarrow$	heineken	0.10	0.34	0.56	-0.39
herring	$\rightarrow$	avocado	0.17	0.34	0.97	-0.02
herring	$\rightarrow$	ham	0.16	0.32	1.09	+0.05
cracker	$\rightarrow$	heineken	0.37	0.75	1.25	+0.34
coke	$\rightarrow$	ice cream	0.22	0.74	2.40	+0.61

The first two rules show rather strong *docs* - yet, assuming the usual coding which implies that we are interested in detecting associations between the '1'-categories of the variables, the rule $X = 0 \rightarrow Y = 1$ is only of minor interest. Thus, the *docs* in the first two plots to the left point towards the "wrong" direction, i.e. the confidence of $\neg X \rightarrow Y$ is a lot higher than $X \rightarrow Y$. The two middle rules hardly show any *doc* at all. This indicates rather weak rules. The last two rules have again strong *docs*, now in the "right" direction. To

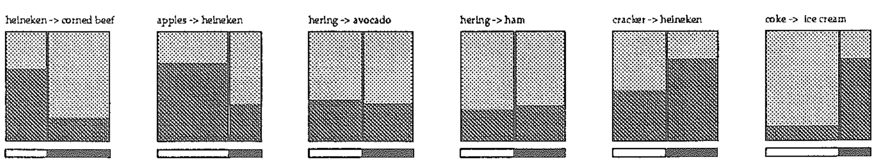


Figure 7: *Six selected association rules - sorted according to their lift and docs.*

decide on the quality of a rule we can perform a test on the *doc*. Since the *doc* can be sen as the difference of proportions ($doc = P(Y \mid X) - P(Y \mid \neg X)$), we can use an approximate Gauss test for testing the equality of two binomial probabilities.

Let $p_1 = conf\,(X \rightarrow Y) = P(Y \mid X)$ and $p_2 = conf\,(\neg X \rightarrow Y) = P(Y \mid \neg X)$ and let n_1 and n_2 be the support of X and $\neg X$. Note that n_1 and n_2 can also be seen as $n \cdot P(X)$ and $n \cdot P(\neg X)$ respectively, where n is the total sample size. With p we denote the weighted mean of p_1 and p_2, i.e. $p = (n_1 p_1 + n_2 p_2)/(n_1 + n_2)$. A test, whether the *doc* is zero or not, is now given by the hypothesis

$$H_0 : doc = 0 \text{ vs. } H_1 : doc > 0$$

and the test statistic

$$T_1 := \frac{p_1 - p_2}{\sqrt{p(1-p)}} \cdot \sqrt{\frac{n_1 n_2}{n_1 + n_2}}$$

which follows a standard normal distribution under the null hypothesis if n_1 and n_2 are sufficiently large. Since we are dealing with large data sets and n_1 exceeds the minimal support, the normal approximation will be valid in most cases.

The test statistic T_1 can be transformed to

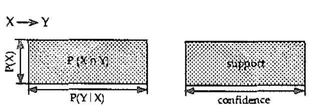
$$T_1^* = \sqrt{n} \cdot \frac{P(X \cap Y) - P(X)P(Y)}{\sqrt{P(X)(1 - P(X)) \cdot P(Y)(1 - P(Y))}},$$

which coincides with the statistic proposed in Piatetsky-Shapiro (1991) and which is identical to the ϕ-coefficient, the square root of the Pearson product-moment correlation. Srikant & Agrawal (1995) claim that only approximately 1% of all the rules found were rejected because of this statistical test. This, however, is strongly dependent both on the specific application and on the choice of minimal confidence and support.

The higher *minconf* and *minsupp* are, the less likely a rule is rejected because of this test. One obvious solution in order to get only rules with a high *doc* is to increase confidence and support. Yet, in most applications we are not interested in the rules with the highest support or confidence - those are quite often known by (data) experts already - but our interest lies in rules at the borderline, which were not known before.

5 Matrix of Association Rules

The basic idea for another approach is to visualise a single association rule not as a square as in Figures 1 and 2 but as a rectangle - this provides us with an additional dimension. Then the use of only one additional colour helps us to gain insight into rules with up to 4-items instead of the 2-items shown in Figure 2:



Support and confidence of each rule $X \to Y$ between two events X and Y can be visualised as a rectangle as follows (see the sketch on the left):

- the *area* of the rectangle is given by *supp* $(X \to Y)$;
- the *height* of the rectangle corresponds to *supp* (X),
  its *width* therefore is $P(X \cap Y)/P(X) = P(Y \mid X)$, the *confidence* of the rule $X \to Y$.

Similar to the visualization techniques introduced above, we plot a matrix of all right hand side items of the rules (response events) vs. the left hand side items (explanatory events). Yet, we want to visualise (by a rectangle as described above) *all* possible combinations between columns and rows. Each column gets the same width, the heights of the rows depend on the corresponding event's probability.

What we are looking for in the plots are relatively large bins (corresponding to a large support) that at the same time have larger widths than heights.

Figure 8 shows a matrix of all 400 theoretically possible 2-item association rules in the SAS assoc data set. The emphasis on the diagonal in this matrix

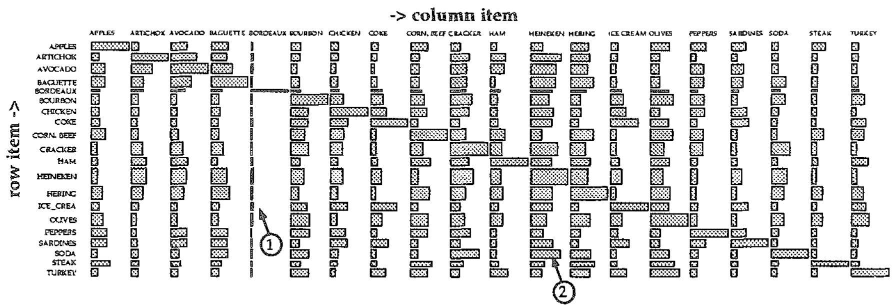


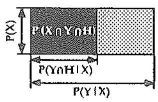
Figure 8: *Matrix of all possible association rules (with two items) from the assocs dataset.*

is obvious since the diagonal represents the support of all the individual items. Actually, the diagonal elements should only be used to measure the frequency of the individual items but not to measure associations. A nice interactive feature would be to allow masking of the diagonal elements. When looking at Figure 8 one immediately notices the unusual behaviour of the fifth column from the left (marked with a circled 1) that corresponds to the item `bordeaux`. This column is almost empty and shows only one bin of significant area. This bin corresponds to the rule `bordeaux` $\rightarrow$ `bordeaux`, indicating, that buying `bordeaux` is not a reaction to buying any of the other items.

The rectangle marked with the circled 2 represents the rule soda $\rightarrow$ heineken. The support of it is fairly large, since the bin's area is large. The confidence of the rule corresponds to the width of the bin and is therefore also relatively high.

5.1 Additional Highlighting

With the means of highlighting we get access to association rules with a higher number of items.



Using two different aspects of the highlighting proportions we are able to look at two different rules at the same time. We can either add the highlighting event to the antecedent part or to the consequent part.
Let H denote some highlighting variable.

1. Adding to the consequence part:
   The highlighting are reflects the joint probability $P(X \cap Y \cap H)$ and since the height represents $P(X)$ it follows that the *absolute width* of the highlighting gives $P(Y \cap H \mid X)$, which is the confidence of the rule $X \to (Y, H)$.

2. Adding to the antecedent part:
   The *proportion* of the highlighted area and the total area of a bin is

$$\frac{P(X \cap Y \cap H)}{P(X \cap Y)} = P(H \mid X{\cap}Y), \text{ which is the confidence of the rule } (X, Y) \to H$$

Let us illustrate this approach with the following two association rules from the SAS Assocs Data,

rule	*conf*	*supp*
sardines → chicken	45.61	12.26
sardines → coke	49.66	14.69

which are used to display the underlying rectangles in Figure 9.

Figure 9: *Rectangles of the association rules sardines → chicken and sardines → coke. Highlighted are transactions including coke & ice cream.*

The highlighting reflects all transactions that include the items coke & ice cream. As described above, we can either concentrate on the absolute number of highlighted cases or on the proportion of highlighted cases vs. the total cases in each of the bins. Each of the views provides us with two more association rules:

1. absolute width of highlighting:

rule $(X \to (Y, H))$	*conf*	*supp*
sardines → chicken & coke & ice cream	39.39	11.59
sardines → coke & ice cream	44.59	13.19

2. proportions of highlighting

rule $((X, Y) \to H)$	*conf*	*supp*
sardines & chicken → coke & ice cream	85.93	11.59
sardines & coke → ice cream	78.91	13.19

What becomes obvious from the example is, that the two different methods for gaining association rules can be applied in different situations: The first method provides rules, which are weaker both with respect to the confidence and support - yet, the statement the new rule makes is stronger than its parent rule ("people buy also H"). Rules derived by using the second method also lose with respect to the support, yet, here the hope is to gain confidence by intensifying the conditions ("people buy Y after buying X and H").

Also note the difference between the rules used in the example. The first rule, *sardines* $\rightarrow$ *chicken*, is independent from the highlighted values (*coke & ice cream*) resulting in a four-item rule. In the second rule the highlighting event overlap in one item with the basic rule *sardines* $\rightarrow$ *coke* which yields to three-item rules only.

Figure 10 shows a highlighted counterpart to Figure 8. The matrix displays all possible association rules with two items and highlighted are all transactions that include coke & ice cream. Since the heights within a row are

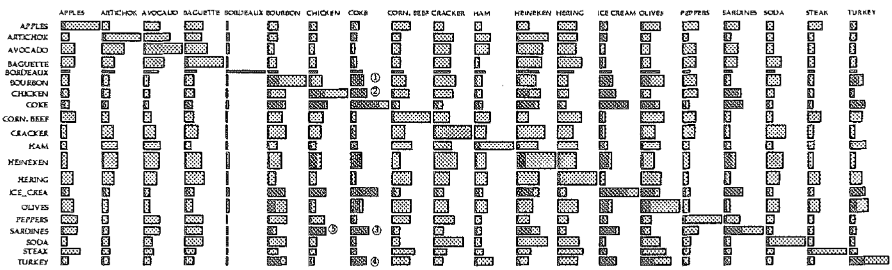


Figure 10: *Matrix of all (2-item) association rules with additional highlighting. Highlighted is the combination coke & ice cream.*

fixed large areas of highlighting correspond to large support rules of the form $X \rightarrow (Y, H)$, e.g.

No	rule	*supp*	*conf*
1	bourbon $\rightarrow$ coke & ice cream	117	$117/403 \approx 0.29$
2	chicken $\rightarrow$ coke & ice cream	123	$123/315 \approx 0.39$
3	sardines $\rightarrow$ coke & ice cream	132	$132/296 \approx 0.45$
4	turkey $\rightarrow$ coke & ice cream	106	$106/283 \approx 0.37$
5	sardines $\rightarrow$ chicken & coke & ice cream	116	$116/296 \approx 0.59$

Rules according to the second pattern, where we regard proportions of highlighting in cells are:

No	rule	*supp*	*conf*
1	bourbon & coke → ice cream	117	$117/140 \approx 0.84$
2	chicken & coke → ice cream	123	$123/139 \approx 0.88$
3	sardines & coke → ice cream	132	$132/147 \approx 0.90$
4	turkey & coke → ice cream	106	$106/119 \approx 0.89$
5	sardines & chicken → coke & ice cream	116	$116/135 \approx 0.86$

Besides the 3- and 4-item association rules we also get insight into the structure among the variables by simply looking for those columns in which the largest highlighting areas occur. Thus, we are able to identify (not necessarily disjoint) clusters of variables. For instance, highlighting the combination **coke & ice cream** shows their link to the variables **bourbon, chicken, heineken, olives, sardines & turkey**

5.2 Matrix of rules and power of rules

We now bring the different pieces of our discussion so far together. To vi-

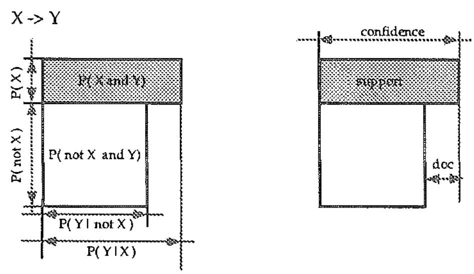


Figure 11: *Schematic Flag plot.*

sualise an association rule together with its *doc* we introduce the concept of a *Flag-plot* (see Figure 11). Beyond the rectangle of the association rule $X \rightarrow Y$, another rectangle is drawn corresponding to the rule $\neg X \rightarrow Y$. The difference in the widths then represents the difference of confidences $P(Y \mid X) - P(Y \mid \neg X) = doc(X \rightarrow Y)$.

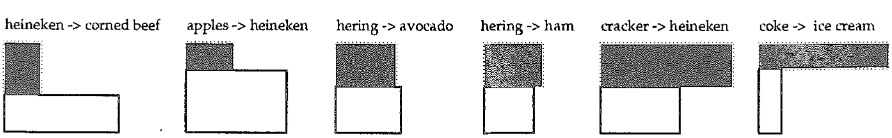


Figure 12: *Six selected association rules - sorted according to doc and lift.*

Figure 12 shows the plots for the six rules already displayed in Figure 7. Figure 13 shows a graphical variation of the matrix in Figure 8. Together with the rectangle corresponding to the association rule between an explanatory

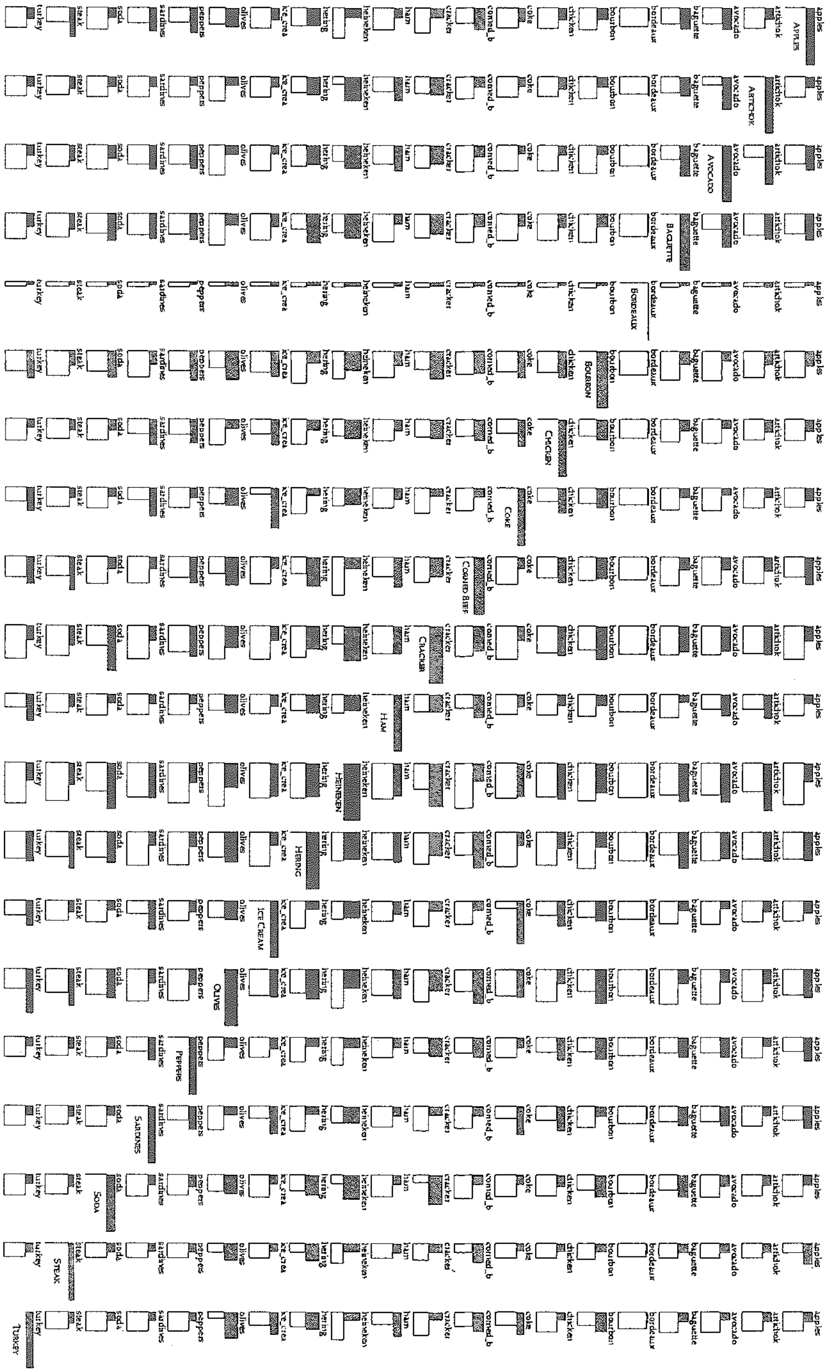


Figure 13: *Association rules of the Assocs data and corresponding docs.*

item and its response also the **negative** rule is shown. From the difference in the rectangles' heights the rule's *doc* can be seen.

6 Conclusions

Association rules are a popular tool in undirected data mining but are not always simple to understand. Powerful filtering mechanisms are needed to reduce the large result sets. For this problem, many techniques have been proposed that filter the most interesting rules (in some, well-defined) sense from the rule set but they neglect the need for explanation of rules. Visual filtering as proposed in this paper can be used in different perspectives. Mosaic plots and , especially, their variant called Doubledecker plots do not only show a single association rule but provide the context of related rules depending on the same contingency table. Thus, the user can better assess the quality of an individual rule. However, this assumes that the vast number of rules generated by the standard association rule algorithms has already been reduced to a manageable size by some automatic filtering tool.

The matrix of association rules is capable to show about 1000 rules on a single page (without displaying the doc) or about 600 rules when displaying the doc. The human eye is well capable to rapidly pick out the largest cells and to detect any particularity in these plots. The conclusions one can draw from these plots give a deeper insight in how (and which) results can be used in solving the business problems one tried to solve using association rules. Further research should focus on combining these visual approaches with other filtering procedures and to work with real data sets form various fields of application. In particular, non-symmetric quantities like $P(Y \mid X) - P(Y)$ will be investigated.

7 Acknowledgement

The authors gratefully acknowledge helpful comments by the referees.

References

Agrawal R., Imielinski T. & Swami A.: *Mining Associations between Sets of Items in Massive Databases.* Proc. of the ACM-SIGMOD 1993 Int'l Conference on Management of Data, Washington D.C., pp. 207 - 216, 1993.

Bayardo R. & Agrawal R.: *Mining the Most Interesting Rules.* In Proc. of the 5th ACM SIGKDD Int'l Conf. on Knowledge Discovery and Data Mining, pp. 145-154, 1999.

Klemettinen M., Mannila H., Ronkainen P., Toivonen H. & Verkamo I.: *Finding interesting rules from large sets of discovered association rules.* In Proceedings of the Third International Conference on Information and Knowledge Management (CIKM'94), ACM Press, pp. 401-407, 1994.

Hofmann H., Siebes A. & Wilhelm A.: *Visualizing Association Rules with Interactive Mosaic Plots.* In Proc. of the 6th ACM SIGKDD Int'l Conf. on Knowledge Discovery and Data Mining, pp. 227-235, 2000.

Hofmann H. & Wilhelm A.: *Validation of Association Rules by Interactive Mosaic Plots.* In J.G. Bethlehem and P.G.M. van der Heijden (eds.), *CompStat 2000 – Proceedings in Computational Statistics*, pp. 499-504, Physica-Verlag, Heidelberg, 2000.

Piatetsky-Shapiro G.: *Discovery, Analysis and Presentation of Strong Rules.* In G. Piatetsky-Shapiro & W. Frawley, *Knowledge Discovery in Databases*, pp.229-248, 1991.

Piatetsky-Shapiro G. & Masand B.: *Estimating Campaign Benefits and Modeling Lift.* In *Proc. of the 5th ACM SIGKDD Int'l Conf. on Knowledge Discovery and Data Mining*, pp.185-193, 1999.

Srikant R. & Agrawal R.: *Mining Generalized Association Rules.* Proc. of the 21st Int'l Conference on Very Large Databases, Zürich, Switzerland, 1995.

Tufte E. (1983), The Visual Display of Quantitative Information. Graphics Press, Chesire Connecticut.

Toivonen H., Klemettinen M., Ronkainen P., Hätönen K. & Mannila H.: *Pruning and grouping discovered association rules.* In MLnet Workshop on Statistics, Machine Learning, and Discovery in Databases, 47-52, Heraklion, Crete, Greece, 1995.