Simultaneous row and column partitioning: The scope of a heuristic approach

Gilbert Ritschard[1] and Djamel A. Zighed[2]

[1] Dept of Econometrics, University of Geneva, CH-1211 Geneva 4, Switzerland
ritschard@themes.unige.ch
[2] ERIC Laboratory, University of Lyon 2, C.P.11 F-69676 Bron Cedex, France
zighed@univ-lyon2.fr

Abstract. This paper is concerned with the determination, in a crosstable, of the simultaneous merging of rows and columns that maximizes the association between the row and column variables. We present an heuristic, first introduced in [21], and discuss its complexity and reliability. The heuristic reduces drastically the complexity of the exhaustive scanning of all possible groupings. Reliability is assessed by means of a series of simulation runs. The outcomes reported show that though the quasi optima often miss the global optima, the discrepancy between them remains extremely low. The extension of the approach to the general multivariate case is also discussed.

1 Introduction

When investigating the relationship between the row and column variables in a cross table, too much empty cells or cells with small frequencies may lead to unreliable results. The usual remedy consists in aggregating categories. This should be done carefully, however, since the merging of categories may affect the strength of the association. It is well known for instance that while merging identically distributed rows or columns does not affect the Pearson or Likelihood ratio Chi-squares (see for example [17] p. 450), it increases generally the value of association measures as illustrated for instance by the simulations reported in [19].

In many settings, the objective is the maximization of the association. For example, in supervised learning we are interested in the best way of using the predictors to discriminate between the values of the response variable. Considering the table that crosses the response variable with the composite variable obtained by crossing all predictors, the problem may be expressed as the search of the simultaneous partitions of the rows and columns that maximize the association between the (row) response variable and the (column) composite predictor. Another example is social sciences, where researchers look for the grouping of the items of surveys that best demonstrates their hypotheses on the role of explanatory factors, the role of social origin on the type of family organization for instance. This paper focuses on this maximization of the association.

Grouping values of variables has to do with categorization and discretization. In this framework, the univariate approach that handles each variable independently has been abundantly studied since the pioneering work of Walter Fisher [9]. The main distinction is here between contextual and non contextual approaches. The former seeks the partitioning that maximizes a contextual criteria like the predictive power of a classifier. See for instance [23] for a survey.

The effect of the groupings on one variable upon the association with a second variable depends obviously on those made on the latter. It is then natural to consider a multidimensional approach that proceeds simultaneously to the aggregation of both variables. The literature about this multidimensional case is less rich. We can mention the related work by Fisher [10, 11] about the optimal grouping of the unknowns and equations of predictive economic models. The simultaneous partitioning of the cases and the variables in a data matrix has been studied among others by Anderberg [1], Bock [5] and Govaert [14]. In [13, 14], Govaert investigates the special case of binary tables. In the framework of contingency tables considered in this paper, the optimal partitioning problem has been studied from different points of view. Benzécri [4] is interested in the partition in a fixed number of groups that maximizes the Pearson Chi-square. A solution to this problem is given in [7] in the form of an iterative heuristic that clusters alternatively the rows and the columns. Gilula and Krieger [12] study how the Pearson Chi-square behaves when the table is reduced by aggregation. Hirotsu [16] and Greenacre [15] are interested in finding the most homogeneous tables. Breiman et al. [6] have studied the joint dichotomization of two variables. Their solution is implemented in their C&RT procedure as the Twoing approach.

The more general multivariate discretization with more than two variables has received some attention from the data mining community. The more significant contributions seem to be those by Douglas et al. [8] who look for the groupings that optimize the predictive accuracy of classifiers, and the non contextual techniques proposed by Monti and Cooper [18] and Bay [2]. In this paper, we assume that we have one target variable and one or several predictors. The general multivariate case is handled simply by taking the composite variable defined by the crossing of the predictors as column variable. The partition of the categories of the composite column variable defines then joint conditional grouping of the predictors.

Our objective is to find both the number of groups and the joint partition of rows and columns of a contingency table that maximizes the association. None of the works cited gives a satisfactory solution to this problem. Some, like those done after Benzécri or those by Breiman et al. assume the number of groups fixed a priori. The others either do not consider the case of contingency tables or consider criteria, homogeneity for example, that are hardly transposable in our setting.

Clearly, the exhaustive scanning of all combinations of partitions of each of the variable is not practicable for large tables. We have thus proposed an heuristic algorithm in [21] that successively looks for the optimal grouping of two row or column categories. The effect of the merging of two categories on the main association measures has been examined analytically in [22]. The purpose of this paper is to further investigate the complexity and the efficacy of the algorithm by completing the preliminary analysis presented in [22, 20].

The formal framework and the notations are defined in Section2. The heuristic is described in Section 3. Section 4 discusses the complexity of the heuristic and compares it with that of the enumeration search of the optimal solution. Section 5 exhibits the outcomes of a series of simulation studies carried out to assess the reliability of the heuristic. Section 6 shortly examines the multivariate case. Finally, Section 7 discusses further developments.

2 Notations and Formal Framework

Let x and y be two variables with respectively r and c different states. Crossing variable x with y generates a contingency table $\mathbf{T}_{r\times c}$ with r rows and c columns.

Let $\theta_{xy} = \theta(\mathbf{T}_{r\times c})$ denote a generic association criterion for table $\mathbf{T}_{r\times c}$. This criterion θ_{xy} may thus be the Cramer v or Kendall τ_b for table $\mathbf{T}_{r\times c}$.

Let P_x be a partition of the values of the row variable x, and P_y a partition of the states of y. Each couple (P_x, P_y) defines then a contingency table $\mathbf{T}(P_x, P_y)$. The optimization problem considered is then the maximization of the association θ_{xy} among the set of couples (P_x, P_y) of partitions

$$\max_{P_x, P_y} \theta\big(\mathbf{T}(P_x, P_y)\big) \ . \tag{1}$$

For ordinal variables, hence for interval or ratio variables, only partitions obtained by merging adjacent categories make sense. We consider then the restricted program

$$\begin{cases} \max\limits_{P_x, P_y} \ \theta\big(\mathbf{T}(P_x, P_y)\big) \\ \text{u.c.} \quad P_x \in \mathcal{A}_x \text{ and } P_y \in \mathcal{A}_y \end{cases} \tag{2}$$

where $\mathcal{A}_x$ and $\mathcal{A}_y$ stand for the sets of partitions obtained by grouping adjacent values of x and y. Letting $\mathcal{P}_x$ and $\mathcal{P}_y$ be the unrestricted sets of partitions, we have for $c, r > 2$, $\mathcal{A}_x \subset \mathcal{P}_x$ and $\mathcal{A}_y \subset \mathcal{P}_y$. Finally, note that ordinal association measures may take negative values. Then, for maximizing the strength of the association, the objective function $\theta\big(\mathbf{T}(P_x, P_y)\big)$ should be the absolute value of the ordinal association measure.

3 The Heuristic

The heuristic, first introduced in [21], is an iterative greedy process that successively merges the two rows or columns that most improve the association criteria

$\theta(\mathbf{T})$. Formally, the configuration (P_x^k, P_y^k) obtained at step k is the solution of

$$\begin{cases} \max\limits_{P_x,P_y} \ \theta\big(\mathbf{T}(P_x,P_y)\big) \\ \text{u.c. } \ P_x = P_x^{(k-1)} \text{ and } P_y \in \mathcal{P}_y^{(k-1)} \\ \qquad \text{or} \\ \qquad P_x \in \mathcal{P}_x^{(k-1)} \text{ and } P_y = P_y^{(k-1)} \end{cases} , \tag{3}$$

where $\mathcal{P}_x^{(k-1)}$ stands for the set of partitions on x resulting from the grouping of two classes of the partition $P_x^{(k-1)}$.

For ordinal variables, $\mathcal{P}_x^{(k-1)}$ and $\mathcal{P}_y^{(k-1)}$ should be replaced by the sets $\mathcal{A}_x^{(k-1)}$ and $\mathcal{A}_y^{(k-1)}$ of partitions resulting from the aggregation of two adjacent elements.

Starting with $\mathbf{T}^0 = \mathbf{T}_{r\times c}$ the table associated to the finest categories of variables x and y, the algorithm successively determines the tables $\mathbf{T}^k, k = 1, 2, \ldots$ corresponding to the partitions solution of (3). The process continues while $\theta(\mathbf{T}^k) \geq \theta(\mathbf{T}^{(k-1)})$ and is stopped when the best grouping of two categories leads to a reduction of the criteria.

The *quasi-optimal grouping* is the couple (P_x^k, P_y^k) solution of (3) at the step k where

$$\theta\big(\mathbf{T}^{(k+1)}\big) - \theta\big(\mathbf{T}^k\big) < 0 \quad \text{and} \quad \theta\big(\mathbf{T}^k\big) - \theta\big(\mathbf{T}^{(k-1)}\big) \geq 0 \ .$$

By convention, we set the value of the association criteria $\theta(\mathbf{T})$ to zero for any table with a single row or column. The algorithm then ends up with such a single value table, if and only if all rows (columns), are equivalently distributed.

4 Complexity

This section compares the complexity of the heuristic to that of the exhaustive exploration of all possible couples (P_x, P_y) of row and column mergings.

For the exhaustive scanning, the number of cases to explore is given by $\#\mathcal{P}_x \#\mathcal{P}_y$, i.e. the number of row partitions times the number of column partitions.

Consider first the case of a nominal variable. The number $B(c) = \#\mathcal{P}$ of possible partitions of its c categories can be computed iteratively by means of Bell's [3] formula $B(c) = \sum_{k=0}^{c-1} \binom{c-1}{k} B(k)$ with $B(0) = 1$. Hence the number of cases to browse is $B(r)B(c)$ in the nominal case, which is for instance about $1.3 \cdot 10^{10}$ for $c = r = 10$.

For ordinal variables, hence for discretization issues, only adjacent groupings are considered. This reduces the number of cases to browse. The number $G(c) = \#\mathcal{A}$ of adjacent groupings of c values is $G(c) = \sum_{k=0}^{c-1} \binom{c-1}{k} = 2^{(c-1)}$. Thus, the number of cases to explore is $G(r)G(c)$ in the ordinal case. For $c = r = 10$ this is for example 262144.

For the heuristic, we can only give the maximal number of couples (P_x, P_y) we may have to scan. The actual number of couples explored depends indeed on

Table 1. Number of configurations explored

	nominal case		ordinal case	
$c = r$	exhaustive	heuristic	exhaustive	heuristic
2	4	4	4	4
3	25	15	16	11
4	225	39	64	22
5	2704	81	256	37
6	41209	146	1024	56
7	769129	239	4096	79
8	17139600	365	16384	106
9	447195609	529	65536	137
10	$1.345 \cdot 10^{10}$	736	262144	172
20	$2.675 \cdot 10^{27}$	6271	$2.749 \cdot 10^{11}$	466
50	$3.449 \cdot 10^{94}$	101676	$3.169 \cdot 10^{29}$	4852
100	$2.264 \cdot 10^{231}$	823351	$4.017 \cdot 10^{59}$	19702

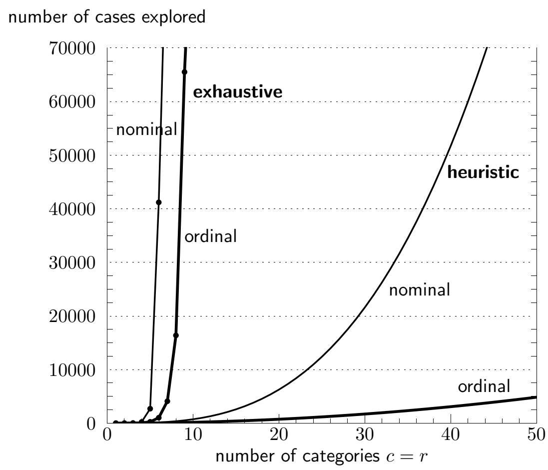

Fig. 1. Complexity versus size of the square table (for heuristic, values reported are upper bounds.)

when the stop criterion is reached. Assuming $c \leq r$, the upper bound is in the nominal case, ,

$$\sum_{i=2}^{c} \left[\binom{i}{2} + \binom{r}{2} \right] + \sum_{j=2}^{r} \binom{j}{2} + 1 = \frac{r(r^2-1) + c(c^2-1)}{6} + \frac{(c-1)r(r-1)}{2} + 1$$

and reads for the ordinal case

$$\frac{(r+c-1)(r+c-2)}{2} + 1 \ .$$

For $c = r = 10$, these bounds are respectively 736 and 172.

Figure 1 and Table 1 shows how the relative efficacy of the heuristic increases with the number of initial categories. The values reported concern square tables. It is worth mentioning that the seemingly exponential increase in the number of cases reported for the heuristic concerns the upper bound. Practically, the effective number of cases browsed will be much lower.

5 Reliability of the Heuristic

The purpose of this section is to assess the reliability of the results provided by the heuristic. A series of simulation studies have been run to investigate two aspects: (i) the number of global optima missed by the heuristic and (ii) how far the quasi-optimum provided by the heuristic is from the global one.

Several association measures have been examined. We report outcomes for the t of Tschuprow, the τ of Goodman and Kruskal and the τ_b of Kendall. Among the measures considered[3] the t of Tschuprow has been retained because it provides the worse scores for both the number of missed optima and the deviations from the global optima. The τ of Goodman and Kruskal has been selected as a representative of the asymmetrical PRE (proportion of reduction in error of prediction) measures. Likewise, the τ_b of Kendall has been selected to represent the ordinal measures.

The comparison between quasi and global optima is done for square tables of size 4, 5 and 6. Above 6, the global optimum can no longer be obtained in a reasonable time.

For the t of Tschuprow and the τ of Goodman and Kruskal, we report respectively in Tables 2 and 3 results for the nominal case as well as for the ordinal case. The τ_b of Kendall being an ordinal measure, Table 4 exhibits only figures for the ordinal case.

For each measure, size and variable type, 200 contingency tables have been randomly generated. Each table was obtained by distributing 10000 cases among its $r \times c$ cells with a random uniform process. This differs from the solution used to generate the results given in [22], which were obtained by distributing the cases with nested conditional uniform distributions: first a random percentage of the cases is attributed to the first row, then a random percentage of the remaining cases is affected to the second row and so on until the last row; the total of each row is then likewise distributed among the columns. The solution retained here generates indeed tables that are closer to the uniform distribution and should therefore exhibit lower association. As will be shown, low association are the less favorable situations for the heuristic. Thus, we can expect the results obtained with this random uniform generating process to provide some upper bounds for the deviations from the global optima.

[3] Simulations have been run for all the measures discussed in [22], i.e. Tschuprow's t, Cramer's v, Goodman & Kruskal's τ, Theil's u, Goodman and Kruskal's γ, Kendall's τ_b and Somers' d. The latter three are ordinal measures and apply therefore only to ordinal variables.

Table 2. Simulations: t of Tschuprow

Tschuprow	nominal			ordinal		
Size	4×4	5×5	6×6	4×4	5×5	6×6
Non zero deviations	39.5%	62.5%	74.5%	23.5%	36%	46.5%
maximum	0.073	0.074	0.077	0.077	0.063	0.108
mean	0.025	0.023	0.028	0.019	0.019	0.012
standard deviation	0.015	0.014	0.016	0.014	0.016	0.015
skewness	0.986	0.979	0.598	1.674	0.972	3.394
With zero deviations						
mean	0.010	0.015	0.021	0.005	0.007	0.006
standard deviation	0.016	0.016	0.018	0.011	0.013	0.012
skewness	1.677	1.062	0.615	3.168	2.211	4.457
Relative deviations						
maximum	0.168	0.198	0.221	0.179	0.194	0.307
mean	0.079	0.077	0.093	0.063	0.066	0.046
Mean initial association	0.260	0.240	0.226	0.263	0.244	0.228
Mean global optimum	0.340	0.316	0.303	0.301	0.275	0.250

Table 3. Simulations: τ of Goodman and Kruskal

G&K τ	nominal			ordinal		
Size	4×4	5×5	6×6	4×4	5×5	6×6
Non zero deviations	5%	6.5%	12%	6%	19%	32.5%
maximum	0.013	0.031	0.029	0.076	0.077	0.059
mean	0.007	0.010	0.008	0.025	0.016	0.013
standard deviation	0.004	0.009	0.009	0.021	0.016	0.012
skewness	-0.308	1.004	1.181	1.107	2.361	1.908
With zero deviations						
mean	0.0004	0.0007	0.0010	0.0015	0.003	0.004
standard deviation	0.0018	0.0033	0.004	0.008	0.009	0.009
skewness	5.323	6.471	5.137	6.685	5.040	3.255
Relative deviations						
maximum	0.142	0.296	0.318	0.420	0.518	0.401
mean	0.062	0.091	0.079	0.216	0.168	0.149
Mean initial association	0.074	0.060	0.048	0.073	0.060	0.050
Mean global optimum	0.148	0.128	0.113	0.118	0.098	0.084

Tables 2 to 4 exhibit, for each series of tables generated, the proportion of optima missed and characteristic values (maximum, mean value, standard deviation, skewness) of the distribution of the deviations between global and quasi optima. Relative deviations, of which the maximum and the mean value are reported, are the ratios between deviations and global optima. The last two rows give respectively the average of the initial values of the criterion and the

Table 4. Simulations: τ_b of Kendall

Kendall τ_b	ordinal		
Size	4×4	5×5	6×6
Non zero deviations	19%	24.5%	32%
maximum	0.596	0.597	0.542
mean	0.235	0.182	0.140
standard deviation	0.195	0.190	0.157
skewness	0.076	0.598	0.652
With zero deviations			
mean	0.045	0.045	0.045
standard deviation	0.125	0.123	0.111
skewness	2.775	2.849	2.445
Relative deviations			
maximum	1.954	1.970	1.982
mean	0.355	0.259	0.074
Mean initial association (abs value)	0.094	0.078	0.064
Mean global optimum (abs value)	0.256	0.236	0.215

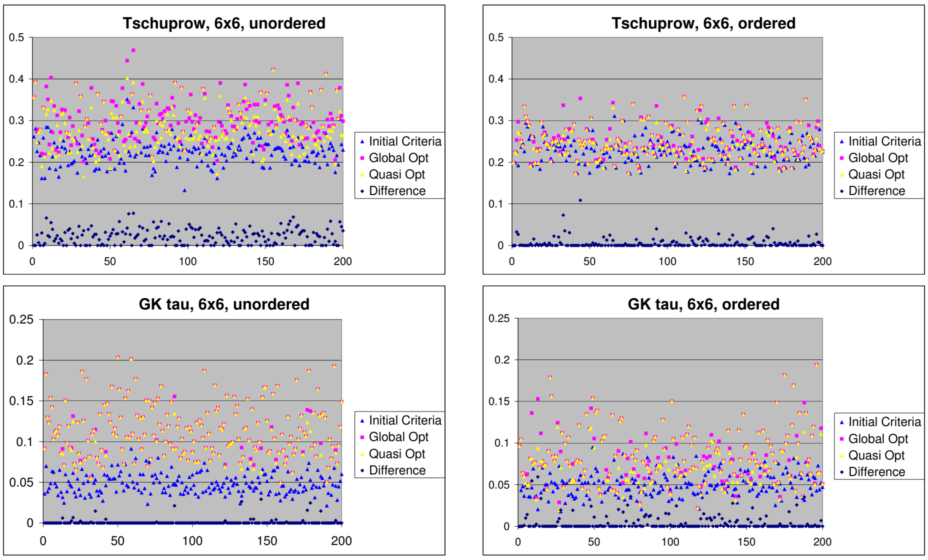

Fig. 2. Initial, quasi and global otpima

mean value of the global optima. In Table 4, these two last figures are means of absolute values of the τ_b's which may be negative.

Additional insight for the Tschuprow's t and Goodman and Kruskal's τ cases is provided by Figures 2 and 3. Figure 2 shows plots of the 200 initial values, quasi optima and global optima for 6×6 cases. Figure 3 plots the 200 deviations against the global optima.

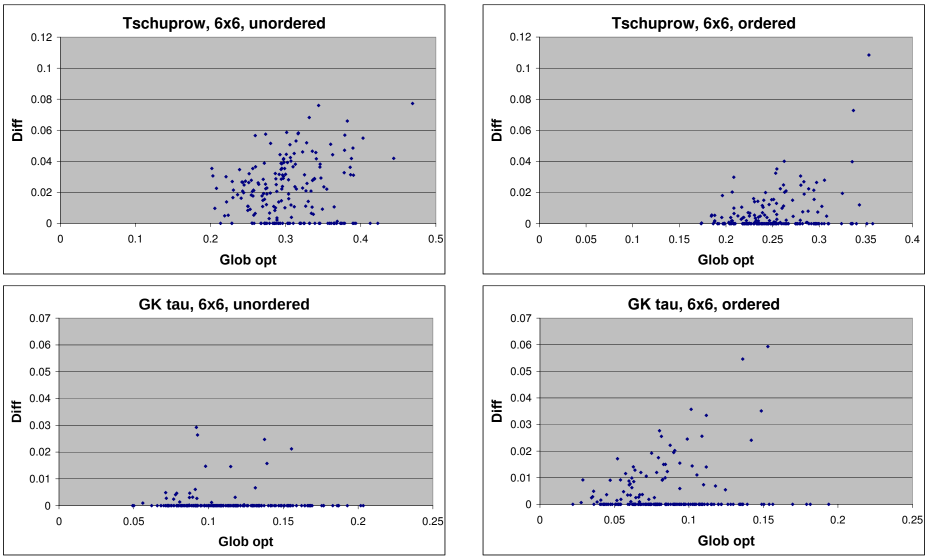

Fig. 3. Deviations versus global otpima

Looking at Tables 2 to 4, we see that the proportion of optima missed by the heuristic is relatively important and tends to increase with the size of the table. The proportion is somewhat lower for PRE measures (the τ of Goodman and Kruskal). This is probably due to the fact that PRE measures cannot be improved by merging values of the predictor (see [22]), which means that the groupings are in this case almost exclusively made on one (the target) variable. Curiously however, the percentages of missed optima are, for PRE measures, larger in the ordinal case than in the nominal one.

This high percentage of missed optima is luckily balanced by the small deviation between the quasi and global optima. The mean value of the non zero deviations is roughly less than half the difference between the initial value of the criterion and the global optimum. In the case of stronger initial associations than those generated here with a uniform random distribution, this ratio becomes largely more favourable, i.e. smaller. The level and dispersion of the non zero deviations seems to remain stable when the size of the table increases. These deviations tend naturally to be larger when the association measures provides larger values. Inversely, the relative deviations take larger values when the association measure tends to zero.

Finally, let us recall that the τ_b of Kendall takes its values in $[-1, 1]$. The deviations may thus exceed the absolute value of the global optimum when the quasi and global optima are of opposite signs. This explains why some maximal relative deviations are greater than one.

Globally, the outcomes of these simulation studies show that the cost in terms of reliability of the heuristic remains moderate when compared with the dramatic increase of performance.

6 A Way to Handle the Multivariate Case

Our approach is intended for the simultaneous partitioning of two variables. There is no straightforward way to extend it to the general multivariate case with more than two variables. On the one hand, it would require the definition of a suitable multivariate association measure, i.e. an index for a multiway crosstable. Measures like the multiple correlation measure the association between one (target) variable and the set of predictors. Hence, they do not measure globally the association between all variables. On the other hand, multiplying the dimensions of the table would dramatically increase the complexity of the heuristic and hence render it unusable.

A solution seems practicable, nevertheless, when we are in presence of one target variable and a set of predictors. In the spirit of the multiple correlation, the multivariate case can in this setting be handled by taking as column variable the composite variable defined by the crossing of the predictors. The optimal grouping of the row target variable and the composite predictor provides then simultaneously the optimal conditional partitions of the predictors and the target variable.

Let us illustrate with an example. The target variable y is dresses quality (high, poor) and the predictors are x_1 the type of dresses (W=women, M=men, C=children) and x_2 the family income (L=low, M=medium, H=high). An optimal solution may then look out as depicted in Table 5:

Table 5. An aggregated multivariate table

	type	W	W	M	C	W	M	M	C	C
quality	income	M	H	H	H	L	L	M	L	M
high			50					10		
poor			5					100		

In this example, we see that medium and low family income are grouped together for men and children while medium and high family income are grouped for women. Likewise, all three categories women, men and children are grouped together for either high income or low income. These demonstrates the conditional aspect of the solutions provided in this multivariate approach.

7 Further Developments

The issue considered is the finding of the optimal way to merge row and column categories in a crosstable. We focused on the maximization of the association.

Other criteria may obviously be considered and should be investigated. For instance, when the data collected in the table are a sample of a larger population, the association computed is an estimate and one should then also care about its standard error. Beside this aspect we are presently working on a strategy to find an optimal aggregation under some constraints. Indeed, following our introductory argument, the primary objective of the reduction of the size of the table is to avoid cells with low frequencies. Therefore, it is worth to be able to maximize for instance the association under a constraint on the minimal cell frequency. On the algorithmic side, we are working on a top-down divisive approach in which, starting from the completely aggregated table we would iteratively split rows or columns. Finally, let us mention that the heuristic discussed has been implemented in the form of a self standing program for the Windows environment. The software is named OCA for Optimal Crosstable Aggregation. It is menu driven and comes with a help and examples files. It is freely downloadable from the first author web site *http://mephisto.unige.ch/stats*.

References

1. Anderberg, M.: *Cluster Analysis for Application.* Academic Press, New York (1973)
2. Bay, S.D.: Multivariate discretization for set mining. *Knowledge and Information Systems* **3** (2001) 491–512
3. Bell, E.T.: The iterated exponential numbers. *Ann. Math.* **39** (1938) 539–557
4. Benzécri, J.P.: *Analyse des données. Tome 2: Analyse des correspondances.* Dunod, Paris (1973)
5. Bock, H.: Simultaneous clustering of objects and variables. In Diday, E., Lebart, L., Pages, J., Tomassone, R., eds.: *Data Analysis and Informatics*, Amsterdam, North Holland (1979) 187–203
6. Breiman, L., Friedman, J.H., Olshen, R.A., Stone, C.J.: *Classification And Regression Trees.* Chapman and Hall, New York (1984)
7. Celeux, G., Diday, E., Govaert, G., Lechevallier, Y., Ralambondrainy, H.: *Classification automatique des données.* Informatique. Dunod, Paris (1988)
8. Dougherty, J., Kohavi, R., Sahami, M.: Supervised and unsupervised discretization of continuous features. In: *Proceedings of the Twelfth International Conference on Machine Learning.* (1995) 194–202
9. Fisher, W.D.: On grouping for maximum of homogeneity. *Journal of the American Statistical Association* **53** (1958) 789–798
10. Fisher, W.D.: Optimal aggregation in multi-equation prediction models. *Econometrica* **30** (1962) 744–769
11. Fisher, W.D.: *Clustering and Aggregation in Economics.* The John Hopkins Press, Baltimore (1969)
12. Gilula, Z., Krieger, A.M.: The decomposability and monotonicity of Pearson's chi-square for collapsed contingency tables with applications. *Journal of the American Statistical Association* **78** (1983) 176–180
13. Govaert, G.: Classification simultanée de tableaux binaires. In Diday, E., Jambu, M., Lebart, L., Pages, J., Tomassone, R., eds.: *Data Analysis and Informatics 3*, Amsterdam, North Holland (1984) 223–236

14. Govaert, G.: Simultaneous clustering of rows and columns. *Control and Cybernetics* **24** (1995) 438–458
15. Greenacre, M.: Clustering the rows and columns of a contingency table. *Journal of Classification* **5** (1988) 39–51
16. Hirotsu, C.: Defining the pattern of association in two-way contingency tables. *Biometrica* **70** (1983) 579–589
17. Jobson, J.D.: *Applied Multivariate Data Analysis*. Volume II: Categorical and Multivariate Methods. Springer-Verlag, New York (1992)
18. Monti, S., Cooper, G.F.: A latent variable model for multivariate discretization. In: *The Seventh International Workshop on Artificial Intelligence and Statistics*. (1999)
19. Olszak, M., Ritschard, G.: The behaviour of nominal and ordinal partial association measures. *The Statistician* **44** (1995) 195–212
20. Ritschard, G.: Performance d'une heuristique d'agrégation optimale bidimensionnelle. *Extraction des connaissances et apprentissage* **1** (2001) 185–196
21. Ritschard, G., Nicoloyannis, N.: Aggregation and association in cross tables. In Zighed, Komorowski, Zytkow, eds.: *Principles of Data Mining and Knowledge Discovery*. Springer-Verlag, Berlin (2000) 593–598
22. Ritschard, G., Zighed, D.A., Nicoloyannis, N.: Maximisation de l'association par regroupement de lignes ou colonnes d'un tableau croisé. *Revue Mathématiques Sciences Humaines* **39** (2001) 81–97
23. Zighed, D.A., Rabaseda, S., Rakotomalala, R., Feschet, F.: Discretization methods in supervised learning. *Encyclopedia of Computer Science and Technology* **40** (1999) 35–50