Agglomerative Hierarchical Multicriteria Clustering Using Decision Rules

V. Batagelj and A. Ferligoj, Ljubljana

Abstract

In a multicriteria clustering problem optimization over more than one criterion is required. The problem can be treated in different ways. In this paper the agglomerative hierarchical method based on decision rules for making decisions under uncertainty is proposed for solving multicriteria clustering problem. An application of proposed approach to Rosenberg and Kim kinship data is presented.

Key words: multicriteria optimization, agglomerative hierarchical clustering methods, decision rules for making decisions under uncertainty.

1 Multicriteria clustering problem

There are clustering problems which can not be appropriately solved with classical clustering methods because they require optimization over more then one criterion. In general, optimal solutions according to each particular criterion are not the same. Because of this the problem arises of how to find the best solution so as to satisfy as much as possible all of the considered criteria.

The usual *clustering problem* (Φ, P) can be formulated as an optimization problem: determine the clustering C^* for which

$$P(C^*) = \min_{C \in \Phi} P(C)$$

where Φ is the set of *feasible clusterings*, C is a clustering and $P : \Phi \to \mathbb{R}$ the (single) *criterion function*.

In a *multicriteria clustering problem* $(\Phi, P_1, P_2, \ldots, P_k)$ we have several criterion functions $P_t, t = 1, \ldots, k$ over the same set of feasible clusterings Φ and our aim is to determine the clustering $C \in \Phi$ in such a way that

$$P_t(C) \to \min, \qquad t = 1, \ldots, k$$

The problem arises of how to find a solution to the problem as good as is possible according to each of the given criteria.

[1]This work was supported in part by the Research Council of Slovenia, Yugoslavia.

Compstat 1990

The multicriteria clustering problem can be treated in different ways: by reduction to a clustering problem with the single criterion obtained as a combination of the given criteria; by clustering with constraints methods where a selected criterion is considered as the clustering criterion and all others determine the constraints; or by direct methods. In [5] two types of direct methods for solving multicriteria clustering problem were proposed: the adapted relocation method, and the adapted agglomerative method. Different elaborations of these two types of methods based on the notion of Pareto-clustering were discussed and compared.

In this paper we elaborate further the agglomerative hierarchical approach based on the criteria compositon.

2 Agglomerative hierarchical approach

We assume that we have k dissimilarity matrices $D^t, t = 1, \ldots, k$, summarizing relevant information on the relationships between the n units, obtained, for example, in k different situations. The problem is to find the best hierarchical solution in a way which satisfies as much as possible all k dissimilarity matrices.

In [5] we proposed two types of adapted agglomerative hierarchical methods for solving the multicriteria clustering problem.

The first approach is to combine the given dissimilarity matrices (at each step) into a composed matrix on which the selection of the nearest pair of clusters is made.

The second approach is to perform the selection step by searching for the *Pareto nearest* pair of clusters. The deficiency of this approach is that the proposed procedure gives several (Pareto) hierarchical solutions. If a smaller set of solutions is desired, additional decision rules or user decisions have to be built into the procedure. Also, there is no single fusion level at each step – there is no simple graphical representation of solution by a dendrogram.

In this paper we follow the first approach. We propose to use at the composition and selection step one of ***decision rules*** (pessimistic, optimistic, Hurwicz, Laplace, Savage) ***for making decisions under uncertainty*** [3, 6]. We obtain the following scheme of the adapted agglomerative method:

> Each unit is a cluster: $C_i = \{X_i\}$, $X_i \in E$, $i = 1, 2, \ldots, n$;
> normalize each dissimilarity matrix $D^t, t = 1, \ldots, k$;
> **repeat** while there exist at least two clusters:
>   determine the nearest pair of clusters C_p in C_q, $d_{p,q} = d(C_p, C_q)$
>   according to a given decision rule;
>   fuse clusters C_p and C_q into a new cluster $C_r = C_p \cup C_q$;
>   replace the clusters C_p and C_q by the cluster C_r;
>   **for each** dissimilarity matrix $D^t, t = 1, \ldots, k$:
>   determine the dissimilarities d^t between the cluster C_r
>   and other clusters according to a given fusion strategy.

The normalization step is not always necessary, especially when dissimilarities are obtained according to the same variables and the same dissimilarity measure but for example on different occasions.

In the pair selection step of the algorithm the decision rules get the form:

Wald's (pessimistic) rule:

$$d_{p,q} = \min_{i,j} \max_{t} d_{i,j}^{t}$$

Optimistic rule:

$$d_{p,q} = \min_{i,j} \min_{t} d_{i,j}^{t}$$

Hurwicz's rule with pessimism index α, $0 \leq \alpha \leq 1$:

$$d_{p,q} = \min_{i,j}(\alpha \max_{t} d_{i,j}^{t} + (1-\alpha) \min_{t} d_{i,j}^{t})$$

Laplace's principle of insufficient reason:

$$d_{p,q} = \frac{1}{k} \min_{i,j} \sum_{t=1}^{k} d_{i,j}^{t}$$

The Savage's minmax regret decision rule

$$d_{p,q} = \min_{i,j} \max_{t} (d_{i,j}^{t} - \min_{t} d_{i,j}^{t})$$

can not be directly applied in this context.

The obtained hierarchical solution can be graphically represented by the dendrogram. The dendrogram levels are the dissimilarities $d_{p,q}$ from the pair selection step.

3 Application: The Rosenberg and Kim data

To illustrate the proposed approach for multicriteria clustering we used the well-known data set on kinship terms [8]. The data are based upon subjective sortings of fifteen kinship terms. Different groups of subjects sorted kinship terms under different sorting instructions with which were obtained six data matrices [1, 62-63]. Two structuring principles could be hypothesized: the first based on kinship and the second based on gender. The data were analyzed by means of hierarchical clustering techniques [8], INDSCAL and INDCLUS models [1], multicriteria relocation methods [4], with algorithm for fitting general graphs to proximity data [7] and other techniques.

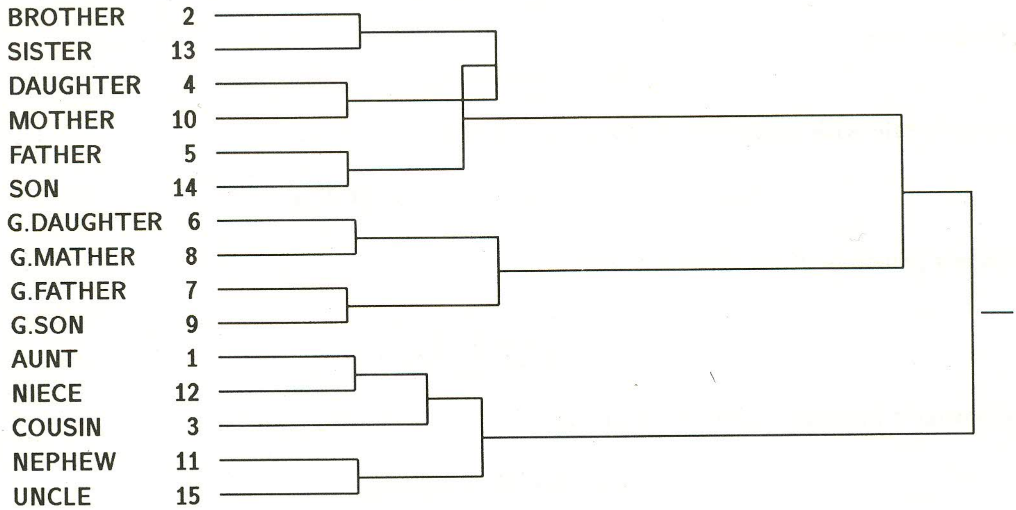
The dendrograms obtained with the proposed agglomerative multicriteria clustering method for Wald's, optimistic, Laplace's and Hurwicz's ($\alpha = 0.25, 0.50$ and 0.75) rules and Ward's criterion function are presented in figures.

All the dendrograms determine the same clustering into three clusters: the first cluster forms the *nuclear family* (mother, father, daughter, son, sister, brother), the second cluster consists of *collaterals* (aunt, uncle, cousin, niece, nephew), and the third cluster are the *grand*parents and *grand*children.

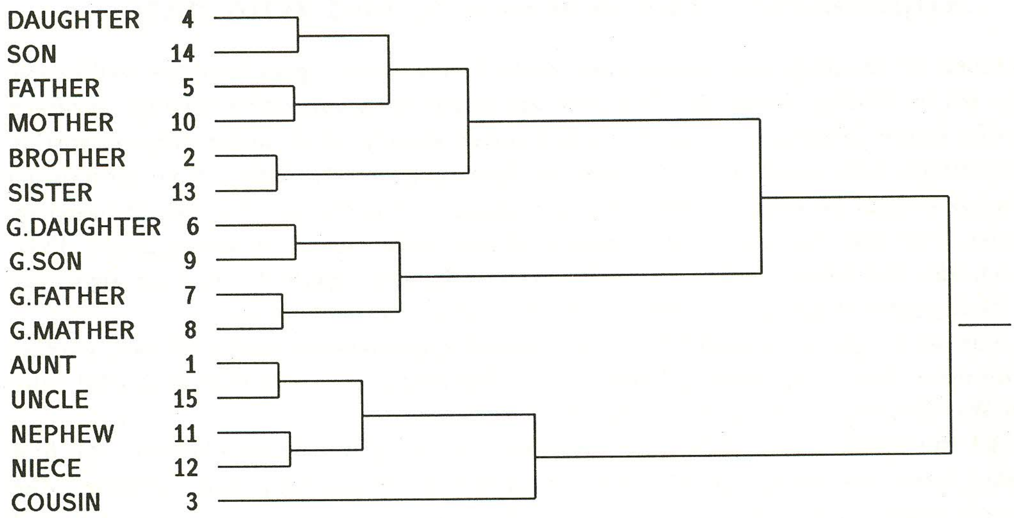
When analyzing each of the six data matrices separately with hierarchical and relocation clustering methods four times the same (previuosly mentioned) clustering was obtained [4].

The dendrograms obtained with four decision rules differ at lower clustering levels: the optimistic and Laplace's dendrograms are very similar – the terms are clustered

CLUSE – plot [0.00, 260.00] May- 1-1990
Wald

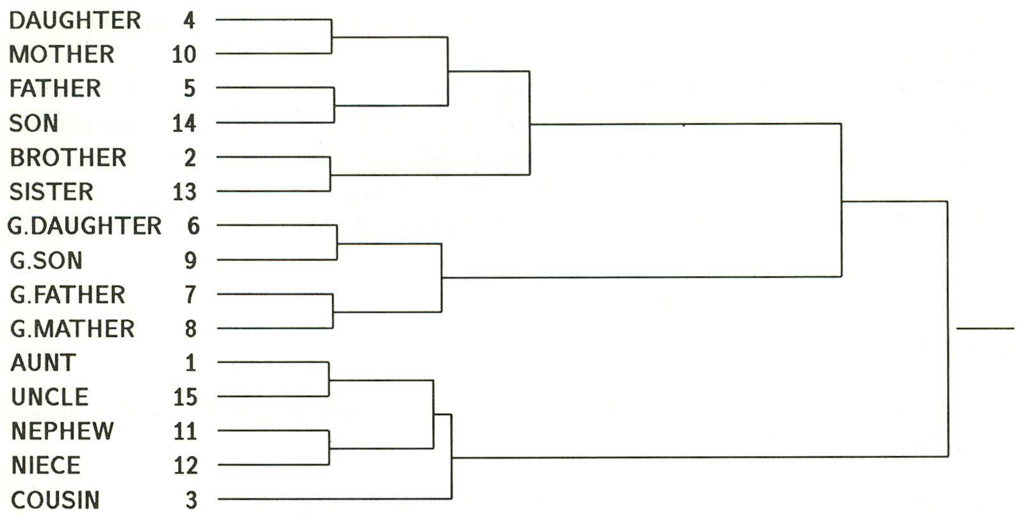


CLUSE – plot [0.00, 130.00] May- 1-1990
Optimist



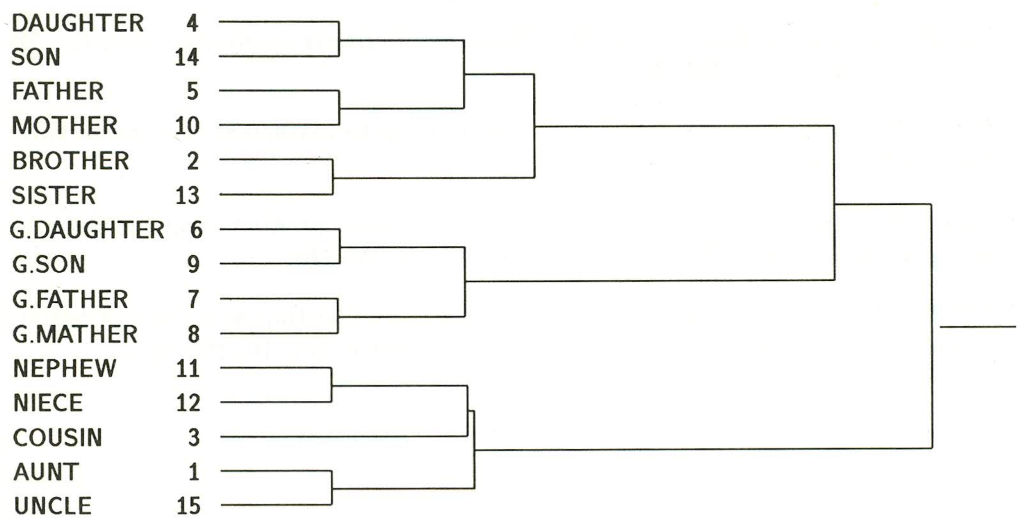
CLUSE – plot [0.00, 200.00] May- 1-1990

Hurwicz 0.5



CLUSE – plot [0.00, 190.00] May- 1-1990

Laplace



on the kinship basis (mother-father, aunt-uncle, brother-sister, ...). Similar is also the clustering produced by the Hurwicz's decision rule with exception of mother and father which are in different clusters. The most different solution was obtained with Wald's rule where the clusters on the lower levels are based on gendre with exception of one cluster (sister, brother).

In summary, the obtained findings are congruent with previous analyses. It is evident that the kinship level has stronger impact to the clustering of kinship terms than gender.

Additional experiments with this approach on different data sets are needed to study the impact of particular decision rule to the corresponding clusterings.

All computations were carried out with the system of clustering programs CLUSE [2].

References

[1] Arabie P., Carroll J.D., and DeSarbo W.S (1987): Three-way Scaling and Clustering. Newbury Park, CA: Sage.

[2] Batagelj V. (1989): CLUSE for IBM PC - Manual. Ljubljana.

[3] Chankong V., and Haimes Y.Y. (1983): Multiobjective Decision Making. New York: North-Holland.

[4] Ferligoj A., and Batagelj V. (1988): Metoda prestavljanj za večkriterijsko razvrščanje v skupine. Statistička revija, 38, 140-154.

[5] Ferligoj A., and Batagelj V. (1989): Direct multicriteria clustering methods. To appear in Journal of Classification.

[6] French S. (1986): Decision Theory, an introduction to mathematics of rationality. New York: John-Wiley.

[7] Klauer K.C., and Carroll J.D. (1989): A mathematical programming approach to fitting general graphs. Journal of Classification, 6, 247-270.

[8] Rosenberg S., and Kim M.P. (1975): The method of sorting as a data-gathering procedure in multivariate research. Multivariate Behaviour, 10, 489-502.