DClust(symbolicDA)

Dynamical clustering based on distance matrix

The aim of the DClust is to partition objects set C into a (fixed) number k of homogenous groups on the basis of the distance matrix. It is an optimization clustering method in which the optimized criterion based on the sum of squares of the distances between objects belonging to the same cluster ($\psi(.)$). The DClust algorithm is as follows (see Diday, Noirhomme-Fraiture [2008], pp. 192):

Step 1. Data set. Symbolic data table containing $c \in C$ objects described by *z* symbolic variables.

Step 2. Initialization. The initial vector of prototypes $q_{(0)} = \{g_1^{(0)},..., g_k^{(0)}\}$ contains random concepts or elements of C.

Step 3. Allocation. An object c_h is assigned to the class $P_i^{(t)}$ if and only if $i = \arg\min\{\psi(c, g_t^{(t-1)}) / l = 1,..., k\}$, k – number of clusters ($P_1,..., P_i,..., P_k$ of the partition *P*).

Step 4. Representation. For $i = 1,..., k$ the prototype g_i^t representing class $P_i^{(t)} \in P^{(t)}$ is the symbolic object of the concept where the index is equal to $h = \arg\min\{\sum_{c_l \in P_i^{(t)}} \psi(c_l, s_m) / m = 1,..., |C|$ then the prototype g_i^t is equal to the symbolic object s_h of the concept c_h.

Step 5. Stopping rule. If $P^{(t)} = P^{(t-1)}$ then stop, else go to allocation step.

References

Bock, H.H., Diday, E. (Eds.) (2000), *Analysis of symbolic data. Explanatory methods for extracting statistical information from complex data*, Springer-Verlag, Berlin.

Diday, E., Noirhomme-Fraiture, M. (Eds.) (2008), *Symbolic Data Analysis with SODAS Software*, John Wiley & Sons, Chichester, pp. 191-204.

Diday, E. (1971), *La methode des Nuees dynamiques*, Revue de Statistique Appliquee, Vol. 19-2, pp. 19-34.

Celeux, G., Diday, E., Govaert, G., Lechevallier, Y., Ralambondrainy, H. (1988), *Classifcation Automatique des Donnees*, Environnement Statistique et Informatique - Dunod, Gauthier-Villards, Paris.