

SClust(symbolicDA)

Dynamical clustering of symbolic data

The aim of the SClust is to partition objects set C into a (fixed) number k of homogenous groups on the basis of symbolic data table. 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 ($\psi(\cdot)$). The DClust algorithm is as follows (see Diday, Noirhomme-Fraiture [2008], pp. 185-186):

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

Step 2. Initialization. Let's $P^{(0)} = \{P_1^{(0)}, \dots, P_k^{(0)}\}$ be the initial random partition of C in k classes.

Step 3. Representation. For $i = 1, \dots, k$ compute a prototype g_i^t as the symbolic object representing the class $P_i \in P^{(t)}$.

Step 4. Allocation. Any concept $c \in C$ is assigned to the class l , if $\psi(c, g_l)$ is minimum:

$$P_i^{(t+1)} = \{c \in C \mid i = \arg \min\{\psi(c, g_l) / l = 1, \dots, k\}.$$

Step 5. Stopping rule. If $P^{(t+1)} = P^{(t)}$ then stop, else go to representation 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. 185-191.
- Verde, R. (2004), *Clustering Methods in Symbolic Data Analysis*, In: D. Banks, L. House, E. R. McMorris, P. Arabie, W. Gaul (Eds.), *Classification, clustering and Data mining applications*, Springer-Verlag, Heidelberg, pp. 299-317.
- 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), *Classification Automatique des Donnees*, Environnement Statistique et Informatique - Dunod, Gauthier-Villards, Paris.