

CEVCLUS algorithm - README

Violaine ANTOINE
violaine.antoine@univ-bpclermont.fr

April 30, 2013

1 Algorithm explanations

The objectif function of CEVCLUS is the following :

$$J_{CEVCLUS}(\mathbf{M}, a, b) = J_{EVCLUS}(\mathbf{M}, a, b) + \xi \frac{1}{2(|\mathcal{M}| + |\mathcal{C}|)} J_{CONST}(\mathbf{M}) \quad (1)$$

where

$$J_{EVCLUS}(\mathbf{M}, a, b) = \frac{1}{\sum_{i < j} d_{ij}} \sum_{i < j} \frac{(aK_{ij} + b - d_{ij})^2}{d_{ij}} \quad (2)$$

and

$$J_{CONST}(\mathbf{M}) = \sum_{(o_i, o_j) \in \mathcal{M}} pl_{i \times j}(\bar{\theta}) + 1 - pl_{i \times j}(\theta) + \sum_{(o_i, o_j) \in \mathcal{C}} pl_{i \times j}(\theta) + 1 - pl_{i \times j}(\bar{\theta}). \quad (3)$$

We consider n objects. The matrix $\mathbf{M} = (m_{ik})$ corresponds to a credal partition and is composed of mass functions. The mass m_{ik} represents the degree of belief that the object \mathbf{x}_i belong the subset A_k , d_{ik} is the distance between \mathbf{x}_i and ω_k , $pl_{i \times j}(\theta)$ is the plausibility that the objects \mathbf{x}_i and \mathbf{x}_j are in the same class and $pl_{i \times j}(\bar{\theta})$ is the plausibility that the two objects are in a different class. The parameter ξ controls the importance given to constraints and K_{ij} represents the degree of conflict between two objects:

$$K_{ij} = 1 - pl_{i \times j}(\theta). \quad (4)$$

Theoretical explanations can be found in [1]. Optimization has been performed as in [2], by using the same parameters.

2 Using CECM script

The CEVCLUS script is a function. Two extra files are provided in order to show how to use the CEVCLUS algorithm. The first one, called *addNewConstraints.m*, randomly selects pairs of objects to create constraints. The second one, referred to as *iris.m*, is a script that loads the Iris dataset and executes on it the CEVCLUS algorithm.

The input arguments of this function are :

- D : a dissimilarity matrix ($n \times p$).
- c : the number of desired clusters.
- link : the constraints matrix ($n \times 3$). The two first columns correspond to the couples of objects (their indexes) and the third column provides the type of constraint : 0 for a Cannot-Link, 1 for a Must-Link.
- Xi : the parameter that enable a tradeoff between the objectif function Jevclus and the constraints.
- Optional :
 - version : choice of subsets for the credal partition
 - 1 : all the 2^c focal elements
 - 2 : singletons + empty set + Ω (value by default)
 - 3 : only singletons

The output arguments of the EVCLUS function are :

- mass : the masses function
- BetP : the pignistic probability
- J : the objective function
- ab : the coefficients in the Stress function J

References

- [1] V. Antoine, B. Quost, M.-H Masson, T. Denoeux, *CEVCLUS: Evidential clustering with instance-level constraints for proximity data*, Soft Computing (to appear), 2013.
- [2] T. Denoeux, M.-H Masson, *EVCLUS: evidential clustering of proximity data*, IEEE Transactions on Systems, Man and Cybernetics: B, vol 34, p 95-109, 2004.