

# CEVCLUS: Constrained evidential clustering of proximity data

Violaine Antoine<sup>1,3</sup> Benjamin Quost<sup>1,3</sup> Marie-Hélène Masson<sup>2,3</sup> Thierry Denœux<sup>1,3</sup>

<sup>1</sup>Université de Technologie de Compiègne

<sup>2</sup>Université de Picardie Jules Verne, IUT de l'Oise

<sup>3</sup>Laboratoire Heudiasyc, UMR CNRS 6599, BP 20529 60205 Compiègne

## Abstract

We present an improved relational clustering method integrating prior information. This new algorithm, entitled CEVCLUS, is based on two concepts: evidential clustering and constraint-based clustering. Evidential clustering uses the Dempster-Shafer theory to assign a mass function to each object. It provides a credal partition, which subsumes the notions of crisp, fuzzy and possibilistic partitions. Constraint-based clustering consists in taking advantage of prior information. Such background knowledge is integrated as an additional term in the cost function. Experiments conducted on synthetic and real data demonstrate the interest of the method, even for unbalanced datasets or non-spherical classes.

**Keywords:** Semi-supervised clustering, pairwise constraints, belief functions, evidence theory, proximity data.

## 1. Introduction

Unsupervised learning analysis refers to a large variety of methods that aim at grouping data into clusters [1]. These algorithms can be distinguished according to the type of input data used. The two most frequent data types are feature vectors in a Euclidean vector space and pairwise proximity measures. The ubiquity of pairwise proximity measures in empirical sciences such as biochemistry, linguistic, psychology and the possibility of feature vectors to be transformed into proximity data make clustering algorithms that can handle proximity data more general than those using feature vectors.

The degree of information provided by a clustering algorithm may be expressed with a crisp, fuzzy, possibilistic or credal partition. Given  $n$  objects to classify in  $c$  classes, a crisp partition aims at grouping each object in one cluster, whereas a fuzzy partition  $U$  classifies each object  $i$  to each cluster  $k$  with a degree of membership  $u_{ik} \in [0, 1]$  such that

$$\sum_{i=1}^c u_{ik} = 1 \quad \forall i \in \{1, \dots, n\}. \quad (1)$$

A possibilistic partition expresses membership values as degrees of possibility. A credal partition,

which extends the concepts of crisp, fuzzy and possibilistic partition, has been recently considered in [2, 3, 4]. This is done by allocating, for each object, a *mass of belief*, not only to single clusters, but also to any subsets of the set of clusters  $\Omega = \{\omega_1, \dots, \omega_c\}$ . The value  $m_i(A)$  such that  $A \subseteq \Omega$  corresponds then to the belief given to  $A$  (and to no more specific subset) regarding the actual cluster of the  $i^{th}$  object. This formulation mainly enables gaining deeper insight into the structure of the data, and also improves the robustness of the algorithm with respect to outliers.

In some situations, extra-information about the data may be available or easy to collect. Making use of these can lead algorithms towards desired solutions. Such algorithms based on background knowledge are known as semi-supervised clustering algorithms. There exists many ways to exploit prior information, at different levels such as the model level [5, 6], the cluster level [7] or the instance level. This last level mainly includes small sets of labelled points. Recently, a weaker form of partial supervision has been considered. Wagstaff [8] proposed to introduce two types of instance-level constraints. A Must-Link constraint specifies that two objects must be in the same class, and a Cannot-Link constraint expresses the fact that two objects should not be in the same class. Such pairwise constraints have been considered and integrated in many unsupervised algorithms such as the hard or the fuzzy c-means (FCM), and have recently become a topic of great interest [9, 10, 11, 12, 13, 14]. They have been incorporated in many different ways, generally by including a penalty term in the objective function [15, 16] or by altering the distances between objects with respect to the constraints [17, 18].

In this paper, we propose to add instance-level constraints in EVCLUS [2], an evidential clustering algorithm dedicated to proximity data. This new algorithm, called CEVCLUS, combines the advantages of adding background knowledge and evidential clustering. The rest of this paper is organized as follows. First, Section 2 recalls the basic background on belief function and evidential clustering with EVCLUS, then Section 3 presents the CEVCLUS algorithm. We initially formulate Must-Link and Cannot-Link constraints in the framework of belief functions to subsequently integrate them in

the evidential clustering scheme. Some results are shown in Section 4. Finally Section 5 concludes and gives some perspectives of the work.

## 2. Background

### 2.1. Belief functions theory

The Dempster-Shafer theory [19, 20], also referred to as belief functions theory, is a mathematical framework for uncertain and imprecise knowledge. The main concepts are presented in this section.

Let  $\Omega$  be a finite set called frame of discernment. A mass function  $m : 2^\Omega \rightarrow [0, 1]$  represents partial knowledge about the actual value taken by a variable  $y$ . This function satisfies

$$\sum_{A \subseteq \Omega} m(A) = 1. \quad (2)$$

The subsets  $A$  such that  $m(A) > 0$  are called the focal elements of  $m$ . The value  $m(A)$  represents the degree of belief that the variable  $y$  belongs to the subset  $A$ , knowing that it has been impossible to assign this belief to a more specific subset of  $A$ . Thus, complete ignorance corresponds to  $m(\Omega) = 1$ . A particular case arises when all focal sets are singletons:  $m$  is then equivalent to a probability distribution. If  $m(\emptyset) > 0$  then  $y$  may not belong to  $\Omega$ . This interpretation is defined as the open-world assumption.

Given a mass function  $m$ , it is possible to define a plausibility function  $pl : 2^\Omega \rightarrow [0, 1]$  and a belief function  $bel : 2^\Omega \rightarrow [0, 1]$  by:

$$pl(A) = \sum_{B \cap A \neq \emptyset} m(B) \quad \forall A \subseteq \Omega, \quad (3)$$

$$bel(A) = \sum_{B \subseteq A, B \neq \emptyset} m(B) \quad \forall A \subseteq \Omega. \quad (4)$$

The quantity  $bel(A)$  represents the total amount of support given to  $A$ , whereas  $pl(A)$  measures the maximal degree of belief that could be given to  $A$ . Those functions are variants of the same information, and one can be retrieved from each other by:

$$pl(A) = 1 - m(\emptyset) - bel(\overline{A}), \quad (5)$$

where  $\overline{A}$  denotes the complement of  $A \subseteq \Omega$ .

The conjunctive rule makes it possible to combine two belief functions defined on the same frame:

$$(m_1 \odot m_2)(A) = \sum_{B \cap C = A} m_1(B) m_2(C) \quad \forall A \subseteq \Omega. \quad (6)$$

The quantity  $(m_1 \odot m_2)(\emptyset)$  is called the degree of conflict between  $m_1$  and  $m_2$ . It denotes the degree of disagreement between the two belief functions, and satisfies

$$K_{12} = (m_1 \odot m_2)(\emptyset) = \sum_{B \cap C = \emptyset} m_1(B) m_2(C). \quad (7)$$

For decision making, the mass function is transformed into a pignistic probability distribution [21]:

$$BetP(\omega) = \sum_{A \in \Omega} \frac{m(A)}{|A|} \quad \forall \omega \in \Omega, \quad (8)$$

where  $|A|$  denotes the cardinality of  $A \subseteq \Omega$ . If  $m(\emptyset) > 0$ , a normalization step has to be performed before carrying out the pignistic transformation. There exists many different methods to execute this step, for example Yager's normalization consists in transferring  $m(\emptyset)$  to  $m(\Omega)$  [22].

### 2.2. Evidential clustering of proximity data

Let us consider  $O = \{o_1, \dots, o_n\}$  a collection of  $n$  objects to classify in a set  $\Omega = \{\omega_1, \dots, \omega_c\}$ . In [2], it was proposed to use the concept of credal partition instead of a crisp or a fuzzy one. Partial knowledge regarding the class membership of an object is represented by a mass function on the set of possible classes. A credal partition of  $O$  is denoted by  $M = (m_1, \dots, m_n)$ .

Let us consider an example of five objects that need to be classified into two classes. A credal partition is shown in Table 1. The classes of the three first objects are known with certainty whereas the class of the fourth object is completely unknown. The fifth object, with all its mass allocated to the empty set, corresponds to an outlier.

$A$	$m_1(A)$	$m_2(A)$	$m_3(A)$	$m_4(A)$	$m_5(A)$
$\emptyset$	0	0	0	0	1
$\{\omega_1\}$	1	1	0	0	0
$\{\omega_2\}$	0	0	1	0	0
$\Omega$	0	0	0	1	0

Table 1: Example of credal partition

Let us suppose that the available data consists of a  $n \times n$  dissimilarity matrix  $\Delta = (\delta_{ij})$  where  $\delta_{ij} = \delta_{ji}$  represents the degree of dissimilarity between two objects  $o_i$  and  $o_j$ . To derive a credal partition from the input dissimilarities, Denœux and Masson [2] proposed the following compatibility condition:

$$\delta_{ij} > \delta_{i'j'} \Rightarrow K_{ij} \geq K_{i'j'}, \quad (9)$$

where  $(o_i, o_j)$  and  $(o_{i'}, o_{j'})$  represent any pairs of object in  $O^2$  and  $K_{ij}$  denotes the degree of conflict between  $m_i$  and  $m_j$ . Thus, the more dissimilar the objects, the higher the conflict between the mass functions. The purpose is then to find a credal partition  $M$  that is the most compatible with  $\Delta$ . This problem is similar to the one addressed by multidimensional scaling algorithms, which aim at finding a configuration of points in a  $p$ -dimensional space such that the distances between points approximate the dissimilarities [23]. The EVCLUS [2] algorithm

minimizes the following stress function:

$$J_{EVCLUS}(M, a, b) = \frac{1}{C} \sum_{i < j} \frac{(aK_{ij} + b - \delta_{ij})^2}{\delta_{ij}}, \quad (10)$$

where  $a$  and  $b$  are two coefficients, and  $C$  is a normalizing constant defined as

$$C = \sum_{i < j} \delta_{ij}. \quad (11)$$

In addition, the credal partition should satisfy the constraints  $m_i(A_k) \geq 0$  for all object  $o_i$  and all subset  $A_k$ , as well as the Equation (2). In order to avoid constrained optimization issues, it is possible to remove the current constraints by using the following parametrization:

$$m_i(A_k) = \frac{\exp(\alpha_{ik})}{2^c \sum_{l=1}^{2^c} \exp(\alpha_{il})}, \quad (12)$$

where  $A_k$ ,  $k = \{1, \dots, 2^c\}$  are the focal sets and the  $\alpha_{ik}$  are the  $(n \times 2^c)$  real parameters representing the credal partition. Thus, the EVCLUS algorithm carries out an iterative optimization of the objective function (10) using a gradient-based procedure.

### 3. EVCLUS with constraints

#### 3.1. Expression of the constraints

Let us consider two objects  $o_i$  and  $o_j$ . We recall that a Must-Link constraint specifies that both objects belong to the same class while a Cannot-Link constraint expresses the fact that the two objects do not belong to the same class. A mass function regarding the joint class membership of both objects may be computed from  $m_i$  and  $m_j$  in the Cartesian product  $\Omega^2 = \Omega \times \Omega$ . This mass function, denoted  $m_{i \times j}$ , is the combination of the vacuous extension of  $m_i$  and  $m_j$ . As shown in [21], we have for all  $A, B \subseteq \Omega$ ,  $A \neq \emptyset, B \neq \emptyset$ :

$$m_{i \times j}(A \times B) = m_i(A) m_j(B), \quad (13)$$

and

$$m_{i \times j}(\emptyset) = m_i(\emptyset) + m_j(\emptyset) - m_i(\emptyset) m_j(\emptyset). \quad (14)$$

In  $\Omega^2$ , a Must-Link corresponds to the subset of  $\Omega$ :  $\theta_{ij} = \{(\omega_1, \omega_1), (\omega_2, \omega_2), \dots, (\omega_c, \omega_c)\}$ , whereas a Cannot-Link corresponds to its complement  $\overline{\theta_{ij}}$ . Then, the plausibilities  $pl_{i \times j}(\theta_{ij})$  and  $pl_{i \times j}(\overline{\theta_{ij}})$  associated with  $m_{i \times j}$  can be determined:

$$pl_{i \times j}(\theta_{ij}) = \sum_{A \cap B \neq \emptyset} m_i(A) m_j(B), \quad (15)$$

and

$$\begin{aligned} pl_{i \times j}(\overline{\theta_{ij}}) &= 1 - m_{i \times j}(\emptyset) - bel_{i \times j}(\theta_{ij}), \\ &= 1 - m_{i \times j}(\emptyset) - \sum_{k=1}^c m_i(\{\omega_k\}) m_j(\{\omega_k\}). \end{aligned} \quad (16)$$

As an example, the plausibilities associated to the credal partition shown in Table 1 are expressed in Table 2. Let us remark that the objects  $o_1$  and  $o_2$  belong to the same class (cf. Table 1); this translates the plausibilities  $pl_{1 \times 2}(\theta_{12}) = 1$  and  $pl_{1 \times 2}(\overline{\theta_{12}}) = 0$ . On the other hand, the objects  $o_1$  and  $o_3$  do not belong to the same class (cf. Table 1), and consequently their joint class plausibilities are  $pl_{1 \times 3}(\theta_{13}) = 0$  and  $pl_{1 \times 3}(\overline{\theta_{13}}) = 1$ . The credal partition does not give any information about the class membership of the object  $o_4$ : its relationship with object  $o_1$  is unknown, and both could either be in the same class or not. Thus,  $pl_{1 \times 4}(\theta_{14}) = pl_{1 \times 4}(\overline{\theta_{14}}) = 1$ . Object  $o_5$  does not belong to  $\Omega$  and represents an outlier. Hence,  $pl_{1 \times 5}(\theta_{15}) = pl_{1 \times 5}(\overline{\theta_{15}}) = 0$ .

Table 2: Plausibilities for the events  $\theta_{ij}$  and  $\overline{\theta_{ij}}$

$F$	$pl_{1 \times 2}(F)$	$pl_{1 \times 3}(F)$	$pl_{1 \times 4}(F)$	$pl_{1 \times 5}(F)$
$\theta_{ij}$	1	0	1	0
$\overline{\theta_{ij}}$	0	1	1	0

In summary, we deduce that the relationship between two objects can be expressed by using the joint plausibilities of both events  $\theta_{ij}$  and  $\overline{\theta_{ij}}$ . Indeed, two objects  $o_i$  and  $o_j$  are surely in the same class if  $pl_{i \times j}(\overline{\theta_{ij}}) = 0$  and  $pl_{i \times j}(\theta_{ij}) = 1$ . Inversely, we assume that two objects  $o_i$  and  $o_j$  are surely in a different class if  $pl_{i \times j}(\theta_{ij}) = 0$  and  $pl_{i \times j}(\overline{\theta_{ij}}) = 1$ .

#### 3.2. Expression of the objective function

In a clustering algorithm, the credal partition is unknown. However, prior information expressed by pairwise constraints may be available. Let us denote by  $\mathcal{M}$  and  $\mathcal{C}$  the sets of Must-Link constraints and Cannot-Link constraints, respectively. A simple formulation of the constraints can be proposed using joint class plausibilities. Indeed, if we know that two objects  $o_i$  and  $o_j$  have a Must-Link constraint, then  $pl_{i \times j}(\overline{\theta_{ij}})$  must be low and  $pl_{i \times j}(\theta_{ij})$  must be high. On the contrary, if we know that two objects  $o_i$  and  $o_j$  have a Cannot-Link constraint, then  $pl_{i \times j}(\theta_{ij})$  must be low and  $pl_{i \times j}(\overline{\theta_{ij}})$  must be close to 1. Let us define  $J_{CONST}$ , the cost of violating pairwise constraints:

$$J_{CONST} = \frac{1}{2(|\mathcal{M}| + |\mathcal{C}|)} (J_{\mathcal{M}} + J_{\mathcal{C}}), \quad (17)$$

$$J_{\mathcal{M}} = \sum_{(o_i, o_j) \in \mathcal{M}} pl_{i \times j}(\overline{\theta_{ij}}) + 1 - pl_{i \times j}(\theta_{ij}), \quad (18)$$

$$J_{\mathcal{C}} = \sum_{(o_i, o_j) \in \mathcal{C}} pl_{i \times j}(\theta_{ij}) + 1 - pl_{i \times j}(\overline{\theta_{ij}}), \quad (19)$$

where  $|\mathcal{M}|$  and  $|\mathcal{C}|$  denote the number of Must-Link constraints and the number of Cannot-Link constraints, respectively.

The CEVCLUS algorithm should compute a credal partition compatible with the input dissimilarities and satisfying as much as possible the instance-level constraints. Thus, we propose to minimize the following objective function:

$$J_{CEVCLUS} = (1 - \xi)J_{EVCLUS} + \xi J_{CONST}, \quad (20)$$

where  $\xi \in [0, 1]$  is a parameter that controls the trade-off between the evidential model and the constraints.

Positivity constraints on mass functions as well as constraint (2) vanish by employing (12). Like in EVCLUS, a gradient based optimization is then applicable to minimize the objective function.

## 4. Results

### 4.1. Methodology

The data sets used in the following experiments include a label for each object. The final partition  $P$  is known and can be compared with a crisp partition  $\hat{P}$  computed by CEVCLUS. A popular measure of agreement between two partitions  $P$  and  $\hat{P}$  is the Rand Index (RI) defined as:

$$RI(P, \hat{P}) = \frac{2(a + b)}{n(n - 1)}, \quad (21)$$

where  $a$  (respectively,  $b$ ) is the number of pairs of objects simultaneously assigned to identical classes (respectively, different classes) in  $P$  and  $\hat{P}$ , and  $n$  is the total number of objects included in the data set.

In order to get a crisp partition  $\hat{P}$ , we transform the final credal partition found with CEVCLUS into a fuzzy partition by using the pignistic transformation. Then we assign each object to the class with maximal pignistic probability.

For each data set in the following experiments, a single trial consists in running 10 times CEVCLUS with different initializations, and retaining the solution with the minimum value of the objective function. Note also that the dissimilarity matrix consists of the matrix of squared Euclidean distances.

### 4.2. Synthetic data set

We first illustrate how the constraint-based algorithm CEVCLUS can be oriented towards a desired solution using a suitable set of constraints. For this purpose, we created an unlabeled synthetic data set. It was generated according to a mixture of four Gaussians in a two-dimensional space. We then defined two labeled data sets by separating the data in two different ways, as shown in Table 3. The data sets are represented in Figures 1 and 2.

In the *Toys1* data set, the two classes are separated by an horizontal boundary. The EVCLUS algorithm applied for this data set gives a credal

Table 3: Construction of the Toys data set

$\mathcal{N}$		Nb	<i>Toys1</i>	<i>Toys2</i>
$\mu$	$\sigma$	objects	classes	classes
(0,0)		100	1	1
(0,7)	$\begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}$	100	1	2
(7,0)		100	2	2
(7,7)		100	2	2

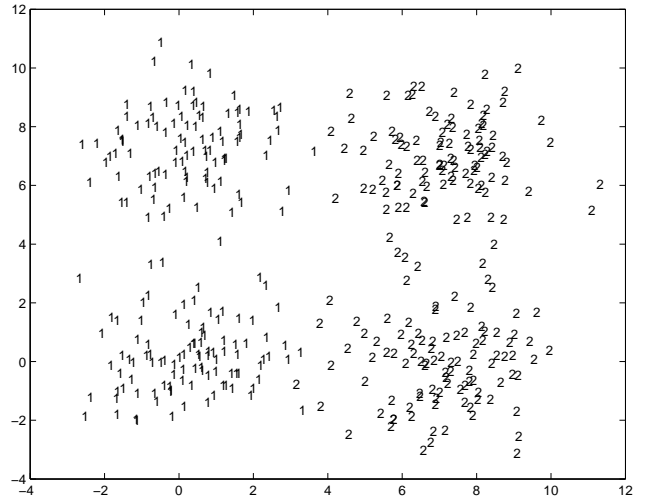


Figure 1: *Toys1* data set.

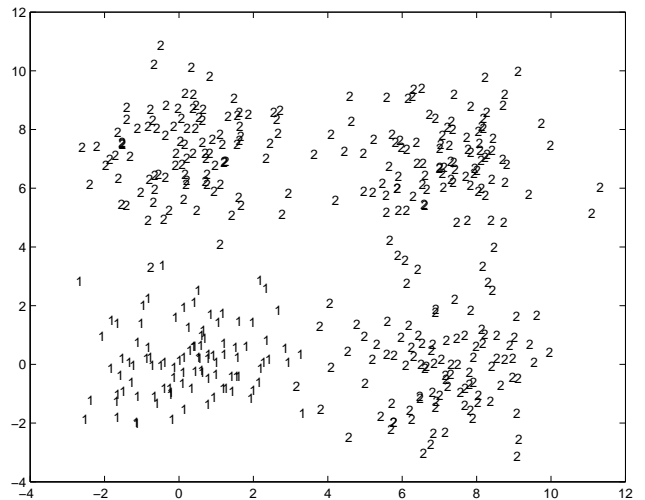


Figure 2: *Toys2* data set.

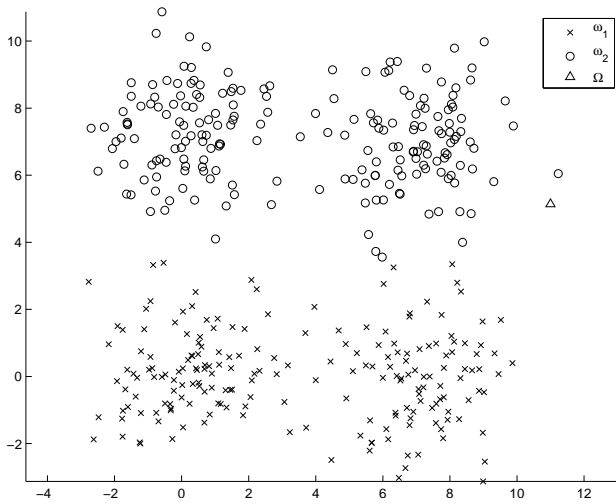


Figure 3: Hard credal partition obtained using EVCLUS.

partition with either vertical or horizontal boundary between the classes, depending on the initialization. The case of an horizontal boundary (cf. Figure 3) does not correspond to the true partition of the *Toys1* data set.

Figure 4 shows that randomly adding 10 constraints makes it possible to guide the algorithm towards the desired solution. Note that constraints were selected by hand so that two constrained objects are not in the same Gaussian. The hard credal partition represent the assignment of each object to the subset  $A \subseteq \Omega$  with the maximal degree of belief. Parameter  $\xi$  was set to 0.2.

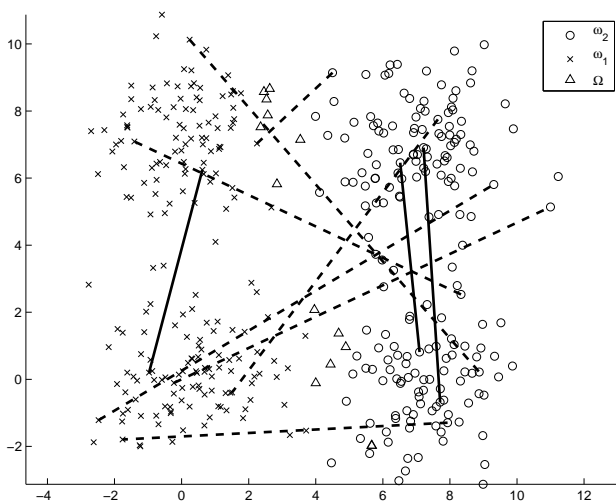


Figure 4: Hard credal partition obtained for *Toys1* using CEVCLUS with 10 constraints. Solid and dashed lines represent Must-Link and Cannot-Link constraints, respectively.

The *Toys2* data set is an unbalanced data set where the boundary between the two classes is non-linear. Applying the EVCLUS algorithm on this

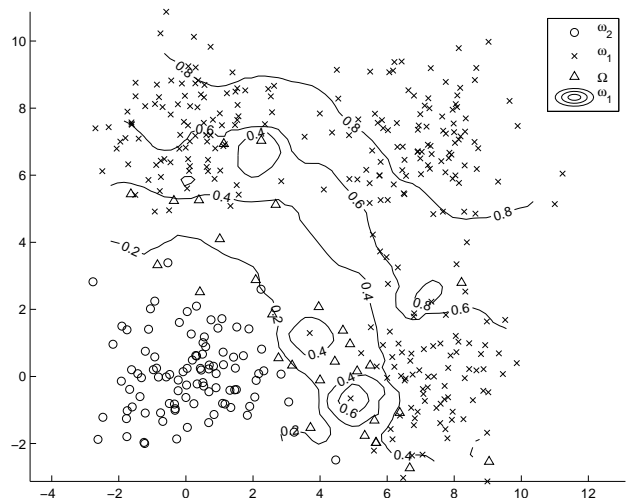


Figure 5: Hard credal partition and contour lines of the mass  $m(\{\omega_1\})$ , obtained for *Toys2* using CEVCLUS with 20 constraints.

data set gives the same solutions that on the *Toys1* data set. Figures 5 and 6 show the hard credal partition given by CEVCLUS with 20 constraints randomly chosen and  $\xi = 0.5$ .

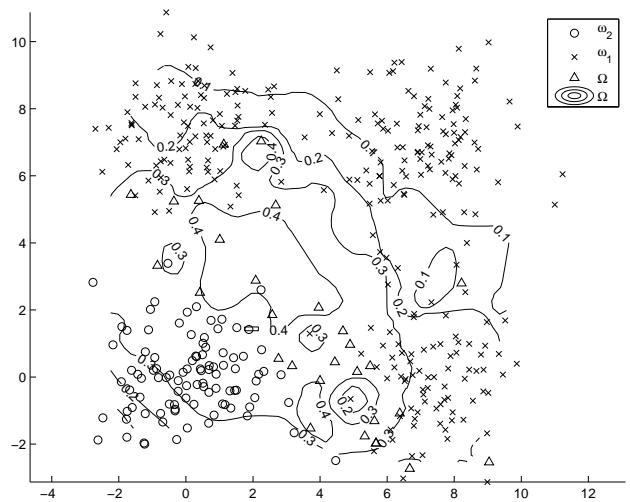


Figure 6: Hard credal partition and contour lines of the mass  $m(\Omega)$ , obtained for *Toys2* using CEVCLUS with 20 constraints.

The degrees of belief of the objects for the sets  $\omega_1$  and  $\Omega$  are also presented with contour lines in those figures. We remark that the constraints allow us to modify the contour lines in such a way that the algorithm finds the desired classes. It can be also observed that the highest degree of belief given to the set of total uncertainty  $\Omega$  is located at the boundary of the two classes. It can be noticed that, unlike previously, the actual partition of the data set *Toys2* is not one of the partitions given by EVCLUS, and is hard to reach as it is far from the initial solutions. Thus, we decided to increase

the importance given to the constraints, by setting  $\xi = 0.5$ .

### 4.3. Iris data set

The Iris data set is composed of three classes that represent different species of Iris: Setosa, Virginica and Versicolor. Each species includes 50 samples. It should be emphasized that the distributions of the classes are not spherical. However, we suppose we do not have this information and we use the Euclidean distances as proximity values.

Figure 7 shows the evolution of the average RI and RI on unconstrained objects (computed over 100 trials) according to the number of constraints. Remark that  $\xi$  is set to 0.2 and five focal elements are selected: the singletons,  $\Omega$  and  $\emptyset$ .

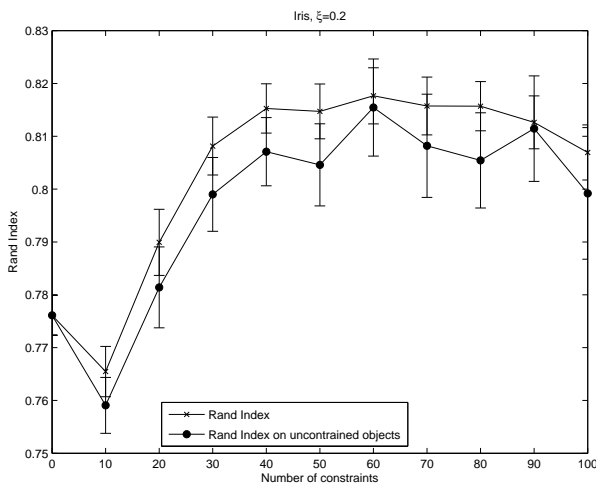


Figure 7: Averaged Rand Index and 95% confidence interval as a function of the number of randomly selected constraints.

Adding constraints not only increases the RI, but also improves the RI computed over unconstrained objects. This shows that including constraints leads the algorithm towards a better partition. Remark that adding a small amount of randomly chosen constraints seems to have a negative impact on the performances, as we can observe on the Figure 7. This happens when constrained objects have been misclassified with a high degree of belief. The same behavior has been noticed in many semi-supervised clustering algorithms and has been discussed in [10, 24].

## 5. Conclusion

We have developed a new algorithm, called CEVCLUS, based on the belief functions theory and that makes it possible to incorporate background knowledge in the classification process. It is an extension of the relational evidential clustering algorithm EVCLUS. CEVCLUS computes a credal partition from a dissimilarity matrix. This notion of credal

partition generalizes those of crisp, fuzzy and possibilistic partitions. Our contribution consists in introducing background knowledge in the clustering process of EVCLUS. This prior information is formulated as pairwise constraints on instances: a Must-Link constraint (respectively, a Cannot-Link constraint) indicates that two objects belong (respectively, do not belong) to the same class.

Experiments show that adding a few number of constraints makes it possible to guide the algorithm towards a desired solution, even for unbalanced and non-spherical classes. The quality of the solution depends on the number of constraints and on the trade-off between the evidential model and the constraints. As a rule, the value of the parameter  $\xi$  should not be close to 1. Furthermore, the value of  $\xi$  should preferably be low when there are few constraints, to prevent constrained objects from being misclassified with a high degree of belief.

Future work may be conducted in the choice of the set of constraints. Indeed, we are interested in adding a small number of very informative constraints. The expected result is to decrease the needed quantity of prior-information and to increase the accuracy of the final partition. Such a method, known as active learning [15, 16], automatically selects pairwise constraints during the clustering process.

## References

- [1] E. Diday and J.C. Simon. Cluster analysis. In K.S. Fu, editor, *Digital Pattern Recognition*, pages 47–94. Springer-Verlag, Berlin, 1976.
- [2] T. Denœux and M.-H. Masson. EVCLUS: evidential clustering of proximity data. *IEEE Trans. Systems, Man and Cybernetics: B*, 34:95–109, 2004.
- [3] M.-H. Masson and T. Denœux. ECM: An evidential version of the fuzzy c-means algorithm. *Pattern Recognition*, 41:1384–1397, 2008.
- [4] M.H. Masson and T. Denœux. RECM: Relational evidential c-means algorithm. *Pattern Recognition Letters*, 30(11):1015–1026, 2009.
- [5] S. Zhong and J. Ghosh. A unified framework for model-based clustering. *The Journal of Machine Learning Research*, 4:1001–1037, 2003.
- [6] D. Gondek and T. Hofmann. Non-redundant data clustering. *Knowledge and Information Systems*, 12(1):1–24, 2007.
- [7] I. Davidson and S. S. Ravi. Agglomerative hierarchical clustering with constraints: Theoretical and empirical results. pages 59–70, Porto, Portugal, 2005. Springer.
- [8] K. Wagstaff, C. Cardie, S. Rogers, and S. Schroedl. Constrained k-means clustering with background knowledge. In *Proceedings of the Eighteenth International Conference on Machine Learning*, pages 577–584, Williamstown, MA, USA, 2001.

- [9] S. Basu, M. Bilenko, and R.J. Mooney. A probabilist framework for semi-supervised clustering. In *Proceedings of the ACM SIGKDD International Conference on knowledge discovery and data mining*, pages 59–68, Seattle, WA, USA, 2004.
- [10] K. Wagstaff. Value, cost, and sharing: Open issues in constrained clustering. In Saso Dzeroski and Jan Struyf, editors, *Knowledge Discovery in Inductive Databases (KDID)*, volume 4747, pages 1–10, Berlin, Germany, September 2006. Springer.
- [11] I. Davidson and S. S. Ravi. Clustering with constraints: Feasibility issues and the k-means algorithm. In *Proceedings of the Fifth SIAM International Conference on Data Mining*, page 138, Newport Beach, CA, USA, 2005. Society for Industrial Mathematics.
- [12] Y. Liu, R. Jin, and A.K. Jain. Boostcluster: boosting clustering by pairwise constraints. In *Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 450–459, San Jose, CA, USA, 2007.
- [13] V. Antoine, B. Quost, M.-H. Masson, and T. Denœux. CECM: Adding pairwise constraints to evidential clustering. In *IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, pages 879–886, Barcelona, Spain, July 2010.
- [14] V. Antoine, B. Quost, M.-H. Masson, and T. Denœux. CECM: Constrained Evidential C-Means algorithm. *Computational Statistics and Data Analysis*, 2011. To appear.
- [15] S. Basu, A. Banerjee, and R.J. Mooney. Active semi-supervision for pairwise constrained clustering. In *Proceedings of the SIAM International Conference on Data Mining*, pages 333–344, Lake Buena Vista, FL, USA, 2004.
- [16] N. Grira, M. Crucianu, and N. Boujemaa. Active semi-supervised fuzzy clustering. *Pattern Recognition*, 41(5):1851–1861, 2008.
- [17] E. Xing, A. Ng, M. Jordan, and S. Russell. Distance metric learning with application to clustering with side-information. pages 521–528. MIT Press, Cambridge, MA, 2002.
- [18] D. Klein, S.D. Kamvar, and C.D. Manning. From instance-level constraints to space-level constraints: Making the most of prior knowledge in data clustering. In *Machine Learning - International Workshop -*, pages 307–314, Sydney, Australia, 2002.
- [19] G. Shafer. *A mathematical theory of evidence*. Princeton university press, Princeton, NJ, 1976.
- [20] A. Dempster. Upper and lower probabilities induced by multivalued mapping. *Annals of Mathematical Statistics*, AMS-38:325–339, 1967.
- [21] P. Smets and R. Kennes. The transferable belief model. *Artificial Intelligence*, 66:191–234, 1994.
- [22] R.R. Yager. On the normalization of fuzzy belief structures. *International Journal of Approximate Reasoning*, 14(2-3):127–153, 1996.
- [23] I. Borg and P. Groenen. Modern multidimensional scaling. *New York: Springer*, 1997.
- [24] I. Davidson S. S., K. Wagstaff, and S. Basu. Measuring constraint-set utility for partitional clustering algorithms. *Lecture Notes in Computer Science*, 4213:115–126, 2006.