

## Kernel Credal C-Means Method Based on Belief Functions

Khawla El Bendadi<sup>1</sup>, Yissam Lakhdar<sup>2</sup>, El Hassan Sbai<sup>2</sup> and Abdourahmane Koita<sup>3</sup>

<sup>1</sup>*Department of physic, Faculty of Sciences, University Moulay Ismail Meknès, Maroc*

<sup>2</sup>*High School of Technology, University Moulay Ismail  
BP 3103 Route d'Agouray, 50006 Toulal, Meknes, Morocco  
IFSTAR, IM, LEPSIS, University Paris-Est  
Paris, France*

<sup>1</sup>*k.elbendadi@gmail.com; lakhdar\_yissam@yahoo.fr, <sup>2</sup>sbaielhassan@yahoo.fr, <sup>3</sup>abdourahmane.koita@ifstar.fr*

### Abstract

*The new approaches based on the credal partition commit objects; not only to singleton clusters but also to meta-clusters; with different masses of beliefs. Among these approaches we find the Credal C-means (CCM). In this article, we introduce a new clustering algorithm called kernel Credal C-means (KCCM), which is a kernel version of CCM algorithm. It is based on the beliefs functions. A second contribution of this paper is an improved CCM algorithm using a non-Euclidean metric unlike CCM which is based on Euclidean distance. In order to show the effectiveness of the proposed method, artificial and real data (vehicle trajectory in a bend) are tested at the end of this paper. A comparison is made to evaluate the results obtained with some methods as Evidential C-means (ECM) and CCM.*

**Keyword:** *Classification, Kernel method, credal partition, belief function, vehicle trajectory, meta-cluster, masses of belief, non-Euclidean metric, C-means*

## 1. Introduction

Clustering is an unsupervised method that divide data into disjoint subsets with high intra-cluster similarity and low inter-cluster similarity. Many clustering algorithms have been proposed in the literature such as the most popular C-means [1] which perform a hard classification and assign each element to a single cluster. We find also fuzzy C-means (FCM) and Noise-Clustering (NC) algorithm [2-3]. These methods work when data are separated and not overlapped.

The belief functions introduced in the Dempster Shafer theory [4] have been widely used to model the uncertain and imprecise information for data classification and information fusion as example [5-6].

The credal partition can be considered as an extension of the various classification concepts, such as hard, fuzzy and possibilistic models [7], because it allows objects to belong not only to singleton clusters, but also to the set of classes called meta-clusters which correspond to the disjunction of several singleton clusters.

Several algorithms based on credal partition have been developed, such as the evidential C-mean [8] which is a method inspired by the FCM algorithm [2]. In ECM, the barycenter of a meta-cluster is obtained by averaging the singletons clusters involved in the meta-cluster. This process is questionable because sometimes the centers of the different clusters are very close, and finally the centers overlap with each other, which is not effective for clustering data.

Belief C-means [9] is proposed to overcome the limitations of the ECM. It was

developed by adopting a different interpretation of the meta-cluster. Based on this method the same authors have developed another method called Credal C-means [10].

The Meta-clusters that are simultaneously close to several clusters play a very important role in characterizing the imprecision of the classification. The objects can be difficult to correctly classify and they should be assigned in the meta-clusters that can positively reflect the degree of imprecision of classification and reduce misclassification rate.

In order to deal with some critical issues in credal clustering, we propose a new algorithm called Kernel Credal C-means (KCCM) based on the concept of credal partition and kernel-induced distance. The kernel methods has been introduced to the fuzzy clustering and evidential classification to perform classification in a feature space of higher dimension.

In this paper, we use the kernel method to construct the nonlinear version of CCM. The basic ideas of KCCM is to map the input data into a feature space with higher dimension using a nonlinear transform and then perform CCM in that feature space. However, for kernel methods, combination or selection of the best kernels from a variety of opportunity is a crucial step. This step requires prior knowledge on the data and the results expected.

For KCCM, the centers of clusters and masses of belief are calculated from the minimization of the objective function.

In the literature, several improvements of methods based on credal partition have been proposed, most are made by replacing the Euclidean metric in the objective function by another metric. An improved version of the CCM algorithm with a non-Euclidean distance is also proposed as a second contribution in this paper.

The performance of the method will be shown and discussed in the article and the results of simulated and real data will be compared with those of the ECM and CCM algorithms. A set of trajectories cornering vehicles (LCPC / Nantes) [11] will be used to show the right behavior of the method.

After a brief recall of ECM and CCM algorithms given in Section 2, we present in detail the Kernel Credal C-means (KCCM) approach and an improved version of CCM with a non-Euclidian metric in Section 3. Some experiments are given in Section 4 to illustrate the effectiveness of KCCM with respect to ECM and CCM approaches with Euclidian and non-Euclidian distances, before concluding this paper in Section 5.

## 2. Background

### 2.1. Theory of Belief Functions

The concept of credal partition is based on the theoretical framework of belief functions and it is interpreted by Smets *et al.* in the transferable belief model [12].

Consider a set of  $c$  classes,  $\Omega = \{\omega_1, \omega_2, \dots, \omega_c\}$  and  $n$  objects to be classified. We propose to represent the partial knowledge on the membership of an object in a class by the mass function  $m$  of  $2^\Omega$  in  $[0,1]$ , verifying

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

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

$$Pl(A) = \sum_{B \cap A \neq \emptyset} m(B) \forall A \subseteq \Omega, \quad (2)$$

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

The first function means the quantity of belief which could be allocated to  $A$ , and the second is the total degree of justified belief in  $A$ .

$$Pl(A) = 1 - bel(A) \quad (4)$$

The transformation of the mass function in pignistic probability is made in order to make a decision for the value of  $y$ .

$$\text{BetP}(\omega) = \sum_{\omega \in A} \frac{m(A)}{|A|} \quad \forall \omega \in \Omega \quad (5)$$

where  $|A|$  represents the cardinal of  $A \subseteq \Omega$ .

Determine a credal partition from data returns to find for each individual  $i$ , the quantities  $m_{ij} = m_i(A_j) (A_j \neq \emptyset, A_j \subseteq \Omega)$  such that  $m_{ij}$  is greater if the distance  $d_{ij}$  between  $i$  and focal element  $A_j$  is small.

## 2.2. Evidential C-means

Evidential C-means [8] is a direct generalization of FCM in the framework of belief functions, and it is based on the credal partition which takes advantage of imprecise classes to express partial knowledge of class memberships.

Let us consider a finite and discrete set of objects  $X$  to be clustered over a given frame of discernment  $\Omega = \{\omega_1, \omega_2, \dots, \omega_c\}$  with  $|\Omega| = c$ , and each data point  $x_i = \{x_{i1}, x_{i2}, \dots, x_{ip}\}$  is a  $p$ -dimension vector.

In ECM, the class membership of an object  $x_i$  is represented by a basic belief assignment (bba)  $m_i(\cdot)$  over the power-set  $2^\Omega$ , which consists of all the subsets of  $\Omega$ . For example, if  $\Omega = \{\omega_1, \omega_2\}$ ,  $2^\Omega = \{\omega_1, \omega_2, \omega_1 \cup \omega_2, \Omega\}$ , and  $2^\Omega$  contains  $2^{|\Omega|}$  elements (clusters). The bba is a function  $m(\cdot)$  from  $2^\Omega$  to  $[0, 1]$  satisfying  $\sum_{A \in 2^\Omega} m(A) = 1$ .

The mass of belief for associating the object  $x_i$  with an element  $A_j$  of  $2^\Omega$ , is determined from the Euclidean distance  $d_{ij}$  between  $x_i$  and barycenter  $\bar{v}_j$ .

$$d_{ij} = \|x_i - \bar{v}_j\|^2 \quad (6)$$

where  $\bar{v}_j$  represent the mean value of the singleton cluster included in  $A_j$ , and is defined mathematically by

$$\bar{v}_j = \frac{1}{|A_j|} \sum_{k=1}^c s_{kj} v_k \quad (7)$$

$$\text{where } s_{kj} = \begin{cases} 1 & \text{if } \omega_k \in A_j; \\ 0 & \text{otherwise;} \end{cases} \quad (8)$$

The objective function is defined by

$$J_{ECM} = \sum_{i=1}^n \sum_{A_j \neq \emptyset} c_j^\alpha m_{ij}^\beta d_{ij}^2 + \sum_{i=1}^n \delta^2 m_{i\emptyset}^\beta \quad (9)$$

constrained on

$$\sum_{i=1}^n m_i(A_j) + m_i(\emptyset) = 1 \quad (10)$$

The parameter  $\alpha$  is a tuning parameter allowing to control the degree of penalization for subsets with high cardinality,  $\beta$  is a weighting exponent and  $\delta$  is an adjustable threshold for detecting the outliers.

Masson and Dencœux are adopted the Lagrange multiplier method to minimize the objective function to find the optimal value for the mass of belief.

The result of minimization of (9) under the constraint (10), adopting the Lagrange method, makes it possible to update the masses  $m$  with the following equation:

$$m_{ij} = \frac{c_j^{\frac{-\alpha}{(\beta-1)}} d_{ij}^{\frac{-2}{(\beta-1)}}}{\sum_{A_k \neq \emptyset} c_k^{\frac{-\alpha}{(\beta-1)}} d_{ik}^{\frac{-2}{(\beta-1)}} + \delta^{\frac{-2}{(\beta-1)}}} \quad (11)$$

## 2.3. Credal C-means

Credal C-means is an alternative version of the Fuzzy C-Means, proposed in order to model and group uncertain and imprecise data. This method is developed to prevent ECM limitations.

In CCM, considers a data set  $X$ , grouped on a frame of discernment  $\Omega = \{\omega_1, \omega_2, \dots, \omega_c\}$  with the corresponding centers  $\{v_1, v_2, \dots, v_c\}$ . the meta-cluster center is defined by the mean value of singleton clusters center's included in this meta-cluster which is obtained by (7) as in ECM.

For CCM algorithm, the clustering of a data set is subjected to constraints in order to allocate an object  $x_i$  to its corresponding cluster (singleton cluster, meta-cluster and outlier cluster).

- If an object  $x_i$  is very close to the center  $v_j$  of a singleton but far from others, it will be assigned to this cluster singleton;
- If an object  $x_i$  is very close to several clusters center simultaneously, it will be difficult to associate it with a singleton cluster, for this the object will be assigned to a meta-cluster. This meta-cluster is represented by the disjunction of singleton cluster;
- If an object  $x_i$  is very far from all the cluster centers for the threshold  $\delta$  which is a parameter that affects the point of overlap, it will be considered outlier.

When the discernment  $\Omega$  contains a large number of elements, the computation of the complexity of the credal partition will be very high which represents a big problem for real applications. In CCM, some meta-cluster that has a high cardinality can be eliminated according to a given threshold  $t_c \in [2, 2^\Omega]$ . This constraint is given by  $S^\Omega = \{A_j \text{ such that } |A_j| \leq t_c\} \subseteq 2^\Omega$ .

The objective function of the CCM method is given by:

$$J_{ccm}(M, V) = \sum_{i=1}^n \sum_{j/A_j \in S^\Omega} m_{ij}^\beta D_{ij}^2 \quad (12)$$

with:

$$D_{ij}^2 = \begin{cases} \delta^2 & ; A_j = \emptyset \\ d_{ij}^2 & ; |A_j| = 1 \\ \frac{\sum_{A_k \in A_j} d_{ik}^2 + \gamma d_{ij}^2}{|A_j| + \gamma} ; |A_j| > 1 \end{cases} \quad (13)$$

where  $M$  is the matrix of mass of belief for all objects and  $V$  is the matrix of the cluster centers.

$J_{ccm}$  must be minimized under the following constraints:

$$\sum_{j/A_j \in S^\Omega} m_{ij} = 1 \quad (14)$$

The setting parameter  $\beta$  is a weighting exponent and  $\gamma$  is the weighting factor of the distance between the object and the center of the meta-cluster. It is determined to find a good compromise between the imprecision and the error rates of classification.

The belief mass matrix  $M = (m_1, m_2, \dots, m_n)$  and the  $V$  cluster center matrix can be obtained by minimizing the objective function under the constraint (14), using the Lagrange multiplier method.

After the update of the masses of beliefs, we can get a new belief mass in detail for different focal elements, including singleton cluster, meta-cluster, and outlier clusters as follows:

$$\left\{ \begin{array}{l} m_{ij} = \frac{\delta^{\beta-1}}{\sum(D)} ; \text{if } A_j = \emptyset \\ m_{ij} = \frac{d_{ij}^{\beta-1}}{\sum(D)} ; |A_j| = 1 \\ m_{ij} = \frac{\sum_{A_k \in A_j} d_{ik}^{\beta-1} + \gamma d_{ij}^{\beta-1}}{\sum(D)} ; \text{if } |A_j| > 1 \end{array} \right. \quad (15)$$

where  $\sum(D)$  is defined by:

$$\sum(D) \triangleq \sum_{A_j \neq \emptyset} \delta^{\beta-1} + \sum_{|A_j|=1} d_{ij}^{\beta-1} + \sum_{|A_j|>1} \frac{\sum_{A_k \in A_j} d_{ik}^{\beta-1} + \gamma d_{ij}^{\beta-1}}{|A_j| + \gamma} \quad (16)$$

The minimization of the objective function with respect to  $v$  is an optimization problem without constraint. The partial derivatives with respect to the center give a linear equation system which can be represented by:

$$B_{c \times n} X_{n \times p} = H_{c \times c} V_{c \times p} \quad (17)$$

where the elements of these matrices are given by:

$$B_{li} \triangleq m_{il}^\beta + \sum_{A_l \in A_j} m_{ij}^\beta \frac{1 + \frac{\gamma}{|A_j|}}{|A_j| + \gamma} \quad (18)$$

$$H_{ll} \triangleq \sum_{i=1}^n m_{il}^\beta + \sum_{i=1}^n \sum_{A_l \in A_j} m_{ij}^\beta \frac{1 + \frac{\gamma}{|A_j|}}{|A_j| + \gamma} \quad (19)$$

$$H_{lq} \triangleq \sum_{i=1}^n \sum_{A_l \in A_k, A_q \in A_k} m_{ik}^\beta \frac{\gamma}{|A_k|^2(|A_k| + \gamma)}; l \neq q \quad (20)$$

### 3. Kernel Credal C-Means Method

#### 3.1. Kernel Method

The kernels based approaches have become popular for several years to solve supervised or unsupervised learning problems. The kernel methods are used to find nonlinear decision functions while fundamentally based on linear methods.

Several methods have also emerged, such as the kernel C-means (KCM) [13], the kernel FCM (KFCM) [14], and the kernel Evidential C-means (KECM) [15]. All these methods use an implicit transformation  $\Phi$  of the data space  $X$  to a higher-dimensional space  $F$ . The kernel  $k$  is defined by:

$$k(x, y) = \Phi(x)^t \Phi(y) \quad (21)$$

We present in the following section a kernel based version for credal C-means algorithm to powerful corresponding nonlinear forms using mercer kernels.

A Gaussian kernel with a width  $\sigma$  is used in this study.

$$k(x, y) = \exp\left(\frac{-\|x-y\|^2}{\sigma^2}\right) \quad (22)$$

The optimal value of  $\sigma$  is problem-dependent and can heavily influence the classification accuracy.

If we confine ourselves to the Gaussian kernel which is used almost exclusively in the literature, then  $k(x, x) = 1$ , and

$$\|\Phi(x_k) - \Phi(v_i)\|^2 = 1 - 2K(x_k, v_i) \quad (23)$$

#### 3.2. The Proposed KCCM Algorithm

To group  $n$  object in  $X = \{x_1, x_2, \dots, x_n\}$  into  $c$  clusters  $\omega_1, \omega_2, \dots, \omega_c$ , with the corresponding center  $\{v_1, v_2, \dots, v_c\}$ , and the credal partition is generalized based on the power set  $2^\Omega$ .

Consider a transformation  $\Phi$  from  $X \subseteq R^p$  to  $F \subseteq R^q$ .  $v_k^\Phi$  the center of class  $\omega_k$  in the space induced by  $\Phi$  and  $V^\Phi$  is the matrix formed by the centers. To construct a kernel version of CCM algorithm, we try to minimize the following objective function:

$$J_{kccm}(M^\Phi, V^\Phi) = \sum_{i=1}^n \sum_{j/A_j \in S^\Omega} m_{ij}^{\Phi\beta} D_{ij}^{\Phi2} \quad (24)$$

$$\text{with: } D_{ij}^{\Phi2} = \begin{cases} \delta^2 & ; A_j = \emptyset \\ d_{ij}^{\Phi2} & ; |A_i| = 1 \\ \frac{\sum_{A_k \in A_j} d_{ik}^{\Phi2} + \gamma d_{ij}^{\Phi2}}{|A_j| + \gamma} & ; |A_i| > 1 \end{cases} \quad (25)$$

where  $D_{ij}^{\Phi2}$  is the distance between the object  $i$  and the focal element  $A_j \subseteq \Omega$  in the space induced by  $\Phi$ .

$$D_{ij}^{\Phi2} \triangleq \|\Phi(x_i) - \Phi(v_j)\|^2 \triangleq K(x_i, x_i) + K(v_j, v_j) - 2K(x_i, v_j) \quad (26)$$

In KCCM, the mass of belief matrix  $M = \{m_1, m_2, \dots, m_n\}$  and the clustering centers matrix  $v_k^\Phi$  can be obtained by minimizing the objective function  $J_{kccm}$  under the

following constraint:

$$\sum_{j/A_j \in S^{\Omega}} m_{ij}^{\phi} = 1 \quad (27)$$

The objective of the minimization of  $J_{KCCM}$  is to calculate the new masses of beliefs.

An optimization process similar to that used in CCM [10] performs minimization of the criterion. This process is based on the method of Lagrange multiplier. In a first step we consider that  $v_k^{\phi}$  is fixed. The Lagrange multiplier  $\lambda_i$  is used to solve the minimization problem with constraint relative to  $M^{\phi}$ , as follow:

$$L(M, \lambda_1, \lambda_2, \dots, \lambda_n) = J_{KCCM}(M^{\phi}, V^{\phi}) - \sum_{i=1}^n \lambda_i \left( \sum_{j/A_j \in S^{\Omega}} m_{ij}^{\phi} - 1 \right) \quad (28)$$

By differentiating the lagrangian  $L$  with respect to  $m_{ij}$  and  $\lambda_i$  setting the derivations to zero, we obtain:

$$\frac{\partial L}{\partial m_{ij}^{\phi}} = \beta m_{ij}^{\phi \beta - 1} D_{ij}^{\phi 2} - \lambda_i = 0 \quad (29)$$

$$\frac{\partial L}{\partial \lambda_i} = \sum_{j/A_j \in S^{\Omega}} m_{ij}^{\phi} - 1 = 0 \quad (30)$$

By developing the equation (29)

$$m_{ij}^{\phi} = \left( \frac{\lambda_i}{\beta} \right)^{\frac{1}{\beta-1}} \left( \frac{1}{D_{ij}^{\phi 2}} \right)^{\frac{1}{\beta-1}} \quad (31)$$

Using (30) and (31)

$$\left( \frac{\lambda_i}{\beta} \right)^{\frac{1}{\beta-1}} = \frac{1}{\sum_{j/A_j \in S^{\Omega}} D_{ij}^{\phi 2}} \quad (32)$$

By replacing (32) in (31)

$$m_{ij}^{\phi} = \frac{D_{ij}^{\phi - \frac{2}{\beta-1}}}{\sum_{A_k \in S^{\Omega}} D_{ik}^{\phi - \frac{2}{\beta-1}}} \quad (33)$$

Using (5), the detailed masses of beliefs can be obtained for different focal elements including singleton cluster, meta-cluster and outlier cluster as follows:

$$\begin{cases} m_{ij}^{\phi} = \frac{\delta^{\frac{-2}{\beta-1}}}{\sum(D)} & ; if A_j = \emptyset \\ m_{ij}^{\phi} = \frac{d_{ij}^{\frac{\phi-2}{\beta-1}}}{\sum(D)} & ; if |A_j| = 1 \\ m_{ij}^{\phi} = \frac{\sum_{A_k \in A_j} d_{ik}^{\phi 2} + \gamma d_{ij}^{\phi 2 \left( \frac{1}{\beta-1} \right)}}{|A_j| + \gamma} & ; if |A_j| > 1 \end{cases} \quad (34)$$

where

$$\sum(D) \triangleq \sum_{A_j = \emptyset} \delta^{-2/\beta-1} + \sum_{|A_j|=1} d_{ij}^{\phi-2/\beta-1} + \sum_{|A_j|>1} \left( \frac{\sum_{A_k \in A_j} d_{ik}^{\phi 2} + \gamma d_{ij}^{\phi 2}}{|A_j| + \gamma} \right)^{1/(\beta-1)} \quad (35)$$

In the second steps,  $M$  is fixed, the minimization of  $J_{KCCM}$  with respect to  $V^{\phi}$  is an unconstrained optimization problem. The partial derivatives of  $J_{KCCM}$  with respect to centers are given by:

$$\frac{\partial J_{KCCM}}{\partial v_l} = \sum_{i=1}^n \sum_{A_l \cap A_j \neq \emptyset} m_{ij}^{\phi \beta} \frac{\partial D_{ij}^{\phi 2}}{\partial v_l} \quad (36)$$

$$\text{with } \frac{\partial D_{ij}^{\phi 2}}{\partial v_l} = \frac{\partial}{\partial v_l} D_{il}^{\phi 2} = \frac{4}{\sigma^2} K(x_i, v_l)(v_l - x_i) \quad (37)$$

$$\frac{\partial D_{ij}^{\phi 2}}{\partial v_l} = \frac{\left[ \frac{4}{\sigma^2} K(x_i, v_l)(v_l - x_i) \right] + \left[ \frac{4\gamma}{|A_j| \sigma^2} K(x_i, v_j)(v_j - x_i) \right]}{|A_j| + \gamma} ; if |A_j| = 1 \quad (38)$$

then

$$\frac{\partial J_{KCCM}}{\partial v_l} = \sum_{i=1}^n m_{il}^{\phi\beta} K(x_i, v_l) (v_l - x_i) + \sum_{i=1}^n \sum_{A_l \in A_j} m_{ij}^{\phi\beta} \frac{K(x_i, v_l)(v_l - x_i) + \frac{\gamma}{|A_j|} K(x_i, v_j)(v_j - x_i)}{|A_l| + \gamma}; \text{ if } |A_j| > 1 \quad (39)$$

By setting these derivatives to zero, we obtain  $c$  linear equation that can be written as:

$$\left[ \sum_{i=1}^n m_{il}^{\phi\beta} K(x_i, v_l) + \sum_{i=1}^n \sum_{A_l \in A_j} m_{ij}^{\phi\beta} \frac{K(x_i, v_l) + \frac{\gamma}{|A_j|} K(x_i, v_j)}{|A_j| + \gamma} \right] x_i = \sum_{i=1}^n \left( m_{il}^{\phi\beta} K(x_i, v_l) \right) v_l + \sum_{i=1}^n \sum_{A_l \in A_j} m_{ij}^{\phi\beta} \left( \frac{K(x_i, v_l) + \frac{\gamma}{|A_j|} K(x_i, v_j)}{|A_j| + \gamma} \right) \quad (40)$$

This linear equation system can be represented by a set of matrices as follows:

$$B_{c \times n}^{\phi} \times X_{n \times p} = H_{c \times c}^{\phi} \times V_{c \times p}^{\phi} \quad (41)$$

Where the elements of these matrices are given by:

$$B_{l \times i}^{\phi} \triangleq m_{il}^{\phi\beta} K(x_i, v_l) + \sum_{A_l \in A_j} m_{ij}^{\phi\beta} \frac{K(x_i, v_l) + \frac{\gamma}{|A_j|} K(x_i, v_j)}{|A_j| + \gamma} \quad (42)$$

$$H_{l \times l}^{\phi} \triangleq \sum_{i=1}^n \left( m_{il}^{\phi\beta} K(x_i, v_l) \right) + \sum_{i=1}^n \sum_{A_l \in A_j} m_{ij}^{\phi\beta} \frac{K(x_i, v_l) + \frac{\gamma}{|A_j|} K(x_i, v_j)}{|A_j| + \gamma} \quad (43)$$

$$H_{l \times q}^{\phi} \triangleq \sum_{i=1}^n \sum_{A_l \in A_k, A_q \in A_k} m_{ik}^{\phi\beta} \frac{\gamma K(x_i, v_j)}{|A_k|^2 (|A_k| + \gamma)}, l \neq q \quad (44)$$

$V^{\phi}$  is the solution of the above linear system (43), solved by using a standard linear system solver as

$$V_{c \times p}^{\phi} = H_{c \times c}^{-\phi} \times B_{c \times n}^{\phi} \times X_{n \times p} \quad (45)$$

As shown in Table1 which represent KCCM algorithm, the masses of belief for the objects are unknown at the beginning, and initialization of the matrix  $M$  of belief mass is generated randomly. This initialization has no effect on the final results for the convergence of clustering methods like FCM, ECM and CCM. In the following section some experiments will be given to show the effectiveness of our approach.

**Table 1. Kernel Credal C-Means (KCCM) Algorithm**

Input	Data set: $X = \{x_1, \dots, x_n\}$ in $\mathbb{R}^p$
Parameters	$c$ : number clusters, $2 \leq c < n$ $t_c$ : meta-cluster threshold (default value $t_c = 2$ ) $\delta > 0$ : outlier threshold $\gamma > 0$ : weight of the distance (default value $\gamma = 1$ ) $\varepsilon > 0$ : termination threshold (default value $\varepsilon = 0.001$ ) $\sigma$ : bandwidth of the kernel.
Initialization	Choose randomly initial mass $M_0$ $t \leftarrow 0$ Repeat $t \leftarrow t + 1$ Compute $B_t^\phi$ and $H_t^\phi$ by (44)–(46); Compute $V_t^\phi$ by solving (47); Compute $M_t^\phi$ using (36); until $\ V_t^\phi - V_{t-1}^\phi\  < \varepsilon$

#### 4. Experimentation

In this section, three experiments is presented to evaluate the performance of KCCM approach. The comparison is made with 4 methods: C-means, Evidential C-means, credal C-means and non-Euclidean version of CCM algorithm (NEuc-CCM).The non-Euclidian metric [17] is defined by:

$$d^2(x_j, v_i) = 1 - \exp(-\rho \|x_j - v_i\|^2) \quad (46)$$

The first experiment, based on simulated data is presented to show the performance of credal classification, by using Gaussian data. The second experiment is an application on road safety which data are collected from the trajectories taken in a bend [11] to show the interest of our method. The different approaches are programmed and tested in Matlab<sup>TM</sup> software.

The credal classification has the advantage of considering the imprecise objects in the box in the imprecise classification and not as misclassified objects (misclassification). In our experiments, the concept of imprecision of the classification and misclassification is introduced to evaluate the performance of our method KCCM.

For this, classification error is calculated by:  $R_e = (N_e / T) * 100\%$ , where  $N_e$  is the number of misclassified objects and  $T$  is the total number of objects under test. The degree of imprecision  $R_{ij}$  is calculated by  $R_{ij} = (N_{ij} / T) * 100\%$ , where  $N_{ij}$  is the number of objects that assigned to the meta-clusters with a cardinality value  $j$ .

The choice of the value of the parameter  $\gamma$  is critical, since it acts on the number of objects in the meta-clusters. Because the right choice for this parameter influence positively on the result of the clustering, we obtain a good compromise between the degree of imprecision and classification error that induces a very good classification rate. In our experiments, several values of the parameter  $\gamma$  are tested.

##### 4.1. Experiment 1

In the first experiment, we consider a particular 3-classes data set with two dimensions that have to be clustered. This data set consists of 600 data points as shown in Figure1. The parameters of the used distribution is shown in Table 1.



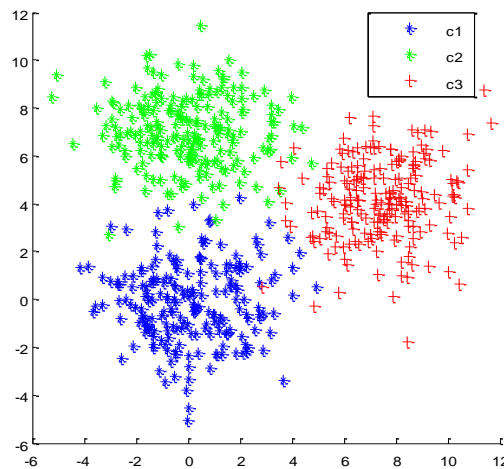
**Table 2. Mean Vector and Covariance Matrix of the Simulated Dataset**

	Mean vector	Covariance matrix	Number of elements
C1	$\begin{bmatrix} 0.0914 \\ -0.0962 \end{bmatrix}$	$\begin{bmatrix} 3.2064 & 0.0843 \\ 0.0843 & 2.7353 \end{bmatrix}$	200
C2	$\begin{bmatrix} -0.0334 \\ 7.0396 \end{bmatrix}$	$\begin{bmatrix} 3.5145 & -0.1365 \\ -0.1365 & 2.2320 \end{bmatrix}$	200
C3	$\begin{bmatrix} 7.3858 \\ 4.1702 \end{bmatrix}$	$\begin{bmatrix} 2.7854 & 0.3730 \\ 0.3730 & 2.8014 \end{bmatrix}$	200

KCCM is applied for the classification of this particular data set and it is compared with the CCM, ECM and non-Euclidian CCM approaches. The outlier threshold used is different from one method to another. For KCCM method we use  $\gamma = 1$ . The three classes represented by the colors red, blue and green are overlapped on their borders as shown in the original data set in Figure1. The classification result obtained by the different methods is shown in Figure 2-5.

ECM produce too much meta-cluster comparing with CCM and KCCM giving three singletons classes  $\{\omega_1\}, \{\omega_2\}, \{\omega_3\}$  as shown in Figure2. The point in the middle of  $\{\omega_1\}$  and  $\{\omega_2\}$ ,  $\{\omega_2\}$  and  $\{\omega_3\}$  are respectively committed to  $\{\omega_1 \cup \omega_2\}$  and  $\{\omega_2 \cup \omega_3\}$  for CCM and KCCM methods because these points are difficult to be correctly classified into a particular class, see Figures 3-5.

In our tests, the value chosen for  $\delta$  differs from one method to another since for NEuc-CCM and KCCM is set to 1. Unlike CCM if this value is small all data points will be committed to the noise class, for this the value of 5 has been chosen.



**Figure 1. Original Data Set**

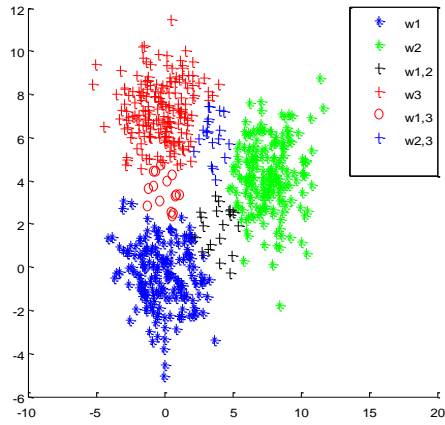


Figure 2. Result with ECM  $\alpha = 1$

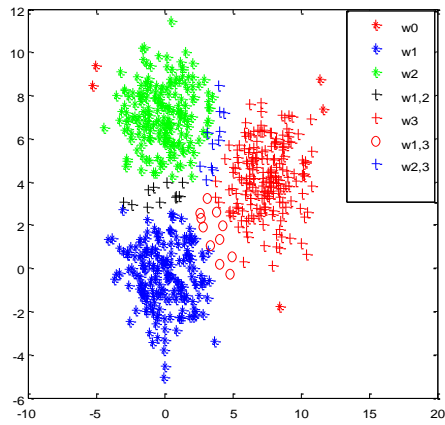


Figure 3. Result with CCM  $\gamma = 1$  and  $\delta = 5$

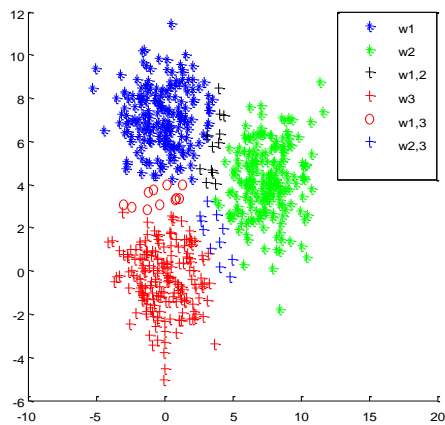
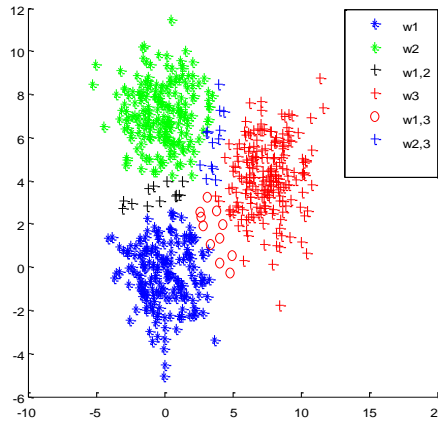


Figure 4. Result with NEuc-CCM  $\gamma = 1$  and  $\delta = 1$



**Figure 5. Result with KCCM  $\gamma = 1$  and  $\delta = 1$**

Despite this choice, as Figure 3 shows, we still find some noisy points, which affects the behavior of the method towards this distribution and increases the error rate.

In Table3, we find the error and imprecision rates for the dataset used in order to compare the pertinence of each method.

The values obtained by the 4 methods show the interest and relevance of our two contributions. Using NEuc-CCM we obtained the smallest value which reaches 5.67% contrary to CCM whose rate equals 6.38%. For KCCM the rate was encouraging if in comparison with CCM and ECM approaches.

**Table 3. Imprecision Rate of Data Set with Different Methods (in %)**

	ECM $\alpha = 1$	CCM $\gamma = 1 \& \delta = 5$	NEuc-CCM $\gamma = 1 \& \delta = 1$	KCCM $\gamma = 1 \& \delta = 1$
$R_e$	0.33	0.67	0.67	0.55
$R_{i2}$	9.83	6.38	5.67	6

We notice that KCCM and CCM produce a lower error rate, it is normal because the objects that difficult to classify are automatically commit to meta-clusters. This shows that the credal classification unlike the hard and fuzzy classification reduces misclassification.

#### 4.2. Experiment 2

In this Experiment, we applied the kernel credal C-means algorithm (KCCM) described in section III to perform the classification of experimental vehicle trajectories. The trajectories used in this study were collected from an instrumented vehicle and a given bend configuration [11]-[18].

The vehicle trajectories is the results of the interaction of the system called Vehicle-Infrastructure-Driver (V-I-D), during an amount of time  $T = [0, t]$  and it can be represented mathematically [17] by:

$$\begin{cases} \Phi : T \rightarrow \mathbb{R}^6 \\ t : \Phi(t) = (x(t), y(t), \frac{dx}{dt}(t), \frac{dy}{dt}(t), \frac{d^2x}{dt^2}(t), \frac{d^2y}{dt^2}(t)) \end{cases} \quad (47)$$

The parameter of function  $\Phi(t)$  represents the six variables, which are the positions  $(x_1, x_2)$ , the speeds  $(v_1, v_2)$  and accelerations  $(\gamma_1, \gamma_2)$ , longitudinal and lateral, respectively, in the Galilean reference. The set of trajectories comprises 232 trajectories. We assume that the observations are representative trajectories practiced in this bend.

To apply KCCM algorithm we used a Gaussian kernel, but he selection of the parameter that defines the kernel (bandwidth for the Gaussian kernel) remains one of the

problems of kernel methods.

The application of KCCM algorithm to the trajectories of vehicles gives four classes as results. These classes represent in the reality the behavior of four types of drivers. The first class represents the family of the slowest trajectories of safe driving. The second class is the family of fast trajectories with safe driving. The third class is the family of the slowest trajectories of sport driving. The fourth class represents the family of the fastest trajectories with sporting driving.

The classification method based on C-means adopted by Koita [11]-[17] allows us to get four classes of stable trajectories. They physically correspond to four behavioral styles of each driver. Table 4 shows the distribution of the number of trajectories observed.

The use of C-means affects all trajectories to a given class. Contrary to CCM method, KCCM produce trajectories belong not only to singleton clusters, but also to meta-clusters.

The obtained results are tested using different values of  $\gamma$ . The parameter for outlier threshold is given by  $\delta = 1$ .

The use of ECM method to classify the trajectories from the bend [19] provided several meta-clusters and we obtained only two classes, after a kernel version of the ECM method have been tested [20], by this we could find the specific four classes desired but we still keeping too much trajectories in meta-clusters.

The use of the KCCM method produces the best results in comparison with CCM. By setting  $\gamma = 0.5$ , we found 25 trajectories not correctly classified that are in meta-clusters, which means that these trajectories have a behavior of two types of drivers. The data belonging to meta-clusters represents a new behavior inspired from at least two behaviors that are established by the experts in the field. These data should be examined to determine to which class belongs.

**Table 4. Clustering Results of Trajectory with Different Methods (Number of Trajectories)**

Methods	Number of trajectories in singleton cluster				Number of trajectories in meta-clusters
	$C_1$	$C_2$	$C_3$	$C_4$	
C-means	35	73	77	47	-
NEuc-CCM $\gamma = 0.5$	35	66	74	41	16
NEuc-CCM $\gamma = 1$	35	59	66	39	33
NEuc-CCM $\gamma = 1,5$	35	55	65	39	38
CCM $\gamma = 0.5$	33	45	61	38	55
CCM $\gamma = 1$	34	44	61	37	56
CCM $\gamma = 1,5$	29	55	60	37	51
KCCM $\gamma = 0.5$	37	62	67	41	25
KCCM $\gamma = 1$	35	59	66	39	33
KCCM $\gamma = 1,5$	35	53	66	38	40

We also notice that NEuc-CCM method gives better results than those obtained with the CCM and the KCCM using trajectories. Because only 16 trajectories have not been assigned to the desired classes. All these tests are compared with the C-means applied by Koita [14].

## 5. Conclusion

The use of our method Kernel Credal C-means algorithm solves the problems in relation to the Euclidean distance by mapping data with nonlinear relationship to appropriate features spaces and we show that KCCM produce a good results using a Gaussian distribution.

We introduced a new based kernel version of credal c-means. The proposed approach is an extension of fuzzy c-means. It is based on belief function theory. KCCM as CCM produce three kinds of clusters: Singleton clusters, meta-clusters and outliers clusters. By introducing meta-clusters, KCCM reduce the misclassification errors. The KCCM method as the CCM approach is able to detect the outliers in the data sets.

Several experiments using the artificial and real data have been presented to evaluate the performance of KCCM with respect to other method. Our results show that KCCM provide a good credal classification result.

The application of our approach on real data applied to road safety; represented by trajectories collected in a bend; gives four classes that represent the behavior of four types of drivers, but our method produce some trajectories that belong to the meta-clusters. These data will be used in future work for the prediction of failed trajectories by using the failure probability.

## References

- [1] G. Ball and D. Hall, "A clustering technique for summarizing multivariate data", *Syst. Res.*, vol. 12, no. 2, (1967), pp. 153-155.
- [2] J. Bezdek, "Pattern Recognition with Fuzzy Objective Function Algorithms", Plenum Press, New York (1981).
- [3] Dave, R. N "Characterization and Detection of Noise in Clustering", *Pattern Recognition Letters*, vol. 12(11), (1991), pp. 657-664.
- [4] G. Shafer. "A Mathematical Theory of Evidence", Princeton University Press, Princeton (1976).
- [5] F. Karem, M. Dhibi and A. Martin, "Combination of supervised and unsupervised classification using the theory of belief functions", *Proceedings of the 2nd International Conference on Belief Functions*, France (2012).
- [6] Z. Liu, J. Dezert, Q. Pan and G. Mercier, "Combination of sources of evidence with different discounting factors based on a new dissimilarity measure", *Decision Support Systems*, vol. 52, no. 1, (2011), pp.133-141.
- [7] R. Krishnapuram and J. Keller, "The possibilistic C-means algorithm: insights and recommendations", *IEEE Trans. Fuzzy Syst.*, vol. 4, no. 3, (1996), pp.385-393.
- [8] M. Masson and T. Denœux, "ECM: An evidential version of the fuzzy c-means algorithm", *Pattern Recognition*, vol. 41, no. 4, (2008), pp. 1384-1397.
- [9] Z. Liu, J. Dezert, G. Mercier and Q. Pan, "Belief C-Means: An extension of Fuzzy C-Means algorithm in belief functions framework", *Pattern Recognition Letters*, vol. 33, no. 3, (2012), pp. 291-300.
- [10] Z. Liu, Q. Pan, J. Dezert and G. Mercier, "Credal c-means clustering method based on belief functions", *Knowledge-Based Systems*, vol. 74, (2015), pp. 119-132.
- [11] A. Koita, D. Daucher and M. Fogli, "New probabilistic approach to estimate vehicle failure trajectories in curve driving", *Probabilistic Engineering Mechanics*, vol. 34, (2013), pp. 73-82.
- [12] P. Smets and R. Kennes, "The transferable belief model", *Artificial Intelligence*, vol. 66, no. 2, (1994), pp. 191-234.
- [13] F. Camastra and A. Verri, "A novel kernel method for clustering", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 5, (2005), pp. 801-805.
- [14] D. Graves and W. Pedrycz, "Kernel-based fuzzy clustering and fuzzy clustering: A comparative experimental study", *Fuzzy Sets and Systems*, vol. 161, no. 4, (2010), pp. 522-543.
- [15] M. H. Masson and T. Denœux, "Evidential c -means in a kernel-induced feature space", *Rencontres francophones sur la Logique Floue et ses Applications*, Lens, France, (2008).
- [16] K. L. Wu and M. S. Yang, "Alternative c-means clustering algorithms", *Pattern recognition*, vol. 35, (2002), pp. 2267-2278.
- [17] D. Daucher and A. Koita, "Modélisation de trajectoires de véhicules en virage par analyse de données expérimentales - application à la sécurité routière", dans *XIXe Congrès français de Mécanique*, Marseille, France, (2009).
- [18] A. Boubezoul, A. Koita and D. Daucher, "Vehicle Trajectories Classification using Support Vectors Machines for Failure Trajectory Prediction", *8th IEEE International Conference on Advances in Computational Tools for Engineering Applications*, Beyrouth, Liban, (2009).

- [19] K. Elbendadi, Y. Lakhdar, E. Sbai and A. Koita, "Classification des trajectoires de véhicules en virage par C-moyenne évidentielle", *Mediterranean Telecommunication*, vol. 4, (2014), pp.62-66.
- [20] K. El Bendadi, Y. Lakhdar, E. Sbai and A. Koita, "Vehicle Trajectories Classification using Kernel Evidential C-means algorithm for Failure Trajectory Prediction", *CESA 6th Multi-conferences on Computational Engineering in Systems Applications IEEE Marrakech, Morocco*, (2015).

## Authors



**Khawla EL Bendadi**, She got a master degree on "computer sciences, computer graphics and imagery" from faculty of sciences, Dhar El Mehraz, university Sidi Mohamed Ben Abdallah, Fez, Morocco, in 2012. Since september 2012, she has been working on his Phd thesis at the university of My Ismail, Meknes, Morocco, with the research team "Control Systems and Information Processing (CSIP)" at higher school of technology of Meknes. His scientific interests include data analysis, Credal Classification, theory of evidence, cluster analysis and machine learning.



**Yissam Lakhdar**, received his Phd thesis at the university of My Ismail, Meknes, Morocco, now, he is attached to the research team "Control Systems and Information Processing (CSIP)" at higher school of technology of Meknes as researcher. His scientific interests include data analysis, artificial intelligence, cluster analysis, image processing and machine learning.



**SBAI El Hassan**, received the Doctorate degree (Doctorat de 3ème cycle) from the university of Sidi Mohammed Ben Abdallah, Fez, Morocco in 1992, and the Doctorate of sciences (Doctorat d'Etat) from the university of My Ismail, Meknes, Morocco in 2001. Since 1993, he is a Professor in automatics and informatics at higher school of technology of Meknes. His main research interest are cluster analysis, pattern recognition, machine learning and computer vision.



**Abdourahmane Koita**, received an engineering degree in mechatronics from UTC (Compiègne) in 2002. In 2011, he obtained a doctorate in random mechanics at the Blaise Pascal University in Clermont Ferrand. He is currently conducting his research activity in the IFSTTAR laboratory. His research focuses on probabilistic modeling and the reliabilistic analysis of the trajectories of multi-body mechanical systems. He has participated in several research projects in the field of road safety.