2013 IEEE International Conference on Systems, Man, and Cybernetics

Credal classification of uncertain data using belief functions

Zhun-ga Liu[a,c], Quan Pan[a], Jean Dezert[b], Gregoire Mercier[c]

a. School of Automation, Northwestern Polytechnical University, Xi'an, China. Email: liuzhunga@gmail.com
b. ONERA - The French Aerospace Lab, F-91761 Palaiseau, France. Email: jean.dezert@onera.fr
c. Telecom Bretagne, CNRS UMR 6285 Lab-STICC/CID, Brest, France, Email: Gregoire.Mercier@telecom-bretagne.eu

***Abstract*—A credal classification rule (CCR) is proposed to deal with the uncertain data under the belief functions framework. CCR allows the objects to belong to not only the specific classes, but also any set of classes (i.e. meta-class) with different masses of belief. In CCR, each specific class is characterized by a class center. Specific class consists of the objects that are very close to the center of this class. A meta-class is used to capture imprecision of the class of the object that is simultaneously close to several centers of specific classes and hard to be correctly committed to a particular class. The belief assignment of the object to a meta-class depends both on the distances to the centers of the specific class included in the meta-class, and on the distance to the meta-class center. Some objects too far from the others will be considered as outliers (noise). CCR provides the robust classification results since it reduces the risk of misclassification errors by increasing the non-specificity. The effectiveness of CCR is illustrated by several experiments using artificial and real data sets.**

***Index Terms*—credal classification, data classification, belief functions, evidence theory.**

I. Introduction

The classical data classification methods are usually developed under the probability framework, like Fuzzy C-means (FCM) [1], artificial neural network (ANN) [2], etc. The data points (objects) can belong to different classes with different probabilities, and the objects are usually assigned to the class with maximum probability. In the classification of uncertain data, several different classes can partly overlap in some cases, and the objects in the overlapped zones are quite hard to be correctly classified into a specific class due to the indistinguishability of these classes for the objects. The probability is not very well adapted to capture such uncertain and imprecise information in the classification problem [3]. Belief function theory [4]–[6], also called evidence theory and Dempster-Shafer theory (DST), is good at dealing with the uncertainty and imprecision of data, and it has been widely applied in the data classification problem [7]–[11].

A recent concept called credal partition have been introduced by Denœux et al in [10] based on belief functions for the unsupervised data clustering, and it can provide a deeper insight to data. Credal partition (classification) allows the objects to belong to not only the singleton (specific) classes, but also the meta-classes (i.e. the disjunction of several specific classes) with different masses of belief, and it can well model the imprecision (ambiguity) of classification. An EVidential CLUStering (EVCLUS) [10] algorithm working with credal partition has been developed for relational data, and Evidential C-Means (ECM) [9] clustering method inspired from FCM [1] has been proposed for the credal partition of object data. Nevertheless, ECM can produce very unreasonable results when the center of the meta-class is close to the specific class, which has been clearly shown and discussed in [8]. In previous related works, we had developed a method called belief c-means (BCM) [8] to overcome the limitation of ECM by introducing an alternative interpretation of the meta-class. BCM purpose was mainly focused on outliers detection, and is rather complicate to use and implement. An evidential EM algorithm [12] has been recently proposed for the parameter estimation in statistical models in the case where data are uncertain and represented by belief functions. Several supervised data classification methods [11] have been also developed based on DST. Particularly, an evidential version of K-nearest neighbors rule (EK-NN) is proposed in [7]. These evidential classifiers work on the specific classes and one extra ignorant class defined by the disjunction of all the specific classes. However, the meta-classes (i.e. the partially ignorant classes), which are very useful and important to explore the imprecision of class of the data, are not considered as available solutions in their classification results.

In this work, we present a new Credal Classification Rule (CCR) taking into account all possible meta-classes for credal classification of uncertain data when the each class can be characterized by the corresponding class center. The object hard to correctly classify should be cautiously committed to the meta-class to reveal its imprecision (ambiguity) degree of the classification and to reduce the risk of misclassification. In CCR, the objects (data points) can be committed with a belief mass function to specific classes, and meta-classes, and eventually to the potential outlier class. The center of each specific class and meta-class is determined at first in the application of CCR. The centers of specific classes can be obtained using the classical way. For example, the class centers produced by FCM can be used in CCR for the unsupervised data classification. When the training data set is available in the supervised data classification problem, the average vector of the training data in each class can be considered as the class center in CCR. The center of meta-class is proposed to be calculated based on the involved specific classes' centers. It

978-1-4799-0652-9/13 $31.00 © 2013 IEEE
DOI 10.1109/SMC.2013.175

IEEE computer society

should have the same (as much as possible similar) distances to each involved specific class center taking into account the data dispersion. In CCR, the specific class contains the objects very close to the center of this class rather than other classes. The object too far from the all the centers will be considered as outlier if the potential outliers (noisy points) are involved. If one object is simultaneously close to several specific classes and it is still quite close to the center of the meta-class defined by the disjunction of these several specific classes, then these specific classes seem undistinguishable for the object according to the distance measure, and it is impossible to make a correct hard assignment of this object to one of these specific classes. Hence, we propose to commit such object to the corresponding meta-class. The interest of credal classification mainly resides in its ability to commit uncertain data points to meta-classes when the available information is insufficient for making specific classification. By doing so, we preserve the robustness of the result and reduce the risk of misclassification errors. Of course the price to pay is the increase of the non-specificity of the classification. In some applications, specially those related to defense and security, it is much more important to maintain such robustness than to provide immediately with high risk a wrong precise classification.

We present in details in Section II the principles of CCR and the mathematical computation of basic belief assignments (bba's) for the credal classification. In Section IV, we present and analyze some simulation results to compare the performances of the CCR with several well-known methods. Conclusions and perspectives will be given in Section V.

II. Credal classification rule (CCR)

A. Basics of belief functions theory

The belief functions have been introduced in 1976 by Shafer in his Mathematical Theory of Evidence known also as Dempster-Shafer theory (DST) [4]–[6]. Let us consider $\Theta = \{\theta_1, \theta_2, \ldots, \theta_h\}$ being a finite discrete set of $h > 1$ mutually exclusive and exhaustive hypotheses θ_i, $i = 1, 2, \ldots, h$. Θ is called the *frame of discernment* of the problem under consideration. The power-set of Θ, 2^Θ, contains all the subsets of Θ. For example, if $\Theta = \{\theta_1, \theta_2, \theta_3\}$, then $2^\Theta = \{\emptyset, \theta_1, \theta_2, \theta_3, \theta_1 \cup \theta_2, \theta_1 \cup \theta_3, \theta_2 \cup \theta_3, \Theta\}$. The union $\theta_i \cup \theta_j = \{\theta_i, \theta_j\}$ (called meta-class) is interpreted as the proposition "the truth value of unknown solution of the problem under concern is either in θ_i, or in θ_j, but θ_i and θ_j are not distinguishable". A basic belief assignment (bba) is a function $m(.)$ from 2^Θ to $[0, 1]$ satisfying $m(\emptyset) = 0$ and $\sum_{A \in 2^\Theta} m(A) = 1$. The *credal partition (classification)* [10] is defined as n-tuple $M = (\mathbf{m_1}, \cdots, \mathbf{m_n})$, where $\mathbf{m}_i$ is the basic belief assignment of the object $\mathbf{x}_i \in X$, $i = 1, \ldots, n$ associated with the different elements of the power-set 2^Θ. The mass of belief of meta-class can well reflect the imprecision (ambiguity) degree of the classification of the uncertain data.

CCR working with credal classification provides a simple and an efficient way to compute the masses of belief associated with the specific classes, the meta-class to characterize the partial imprecision of class, and outlier class. In CCR, it mainly consists of two steps: 1) the determination of the centers of specific and meta-class 2) construction of bba's based on the distance between the object and each class center.

B. Determination of the center of class

Let us consider a data set $X = \{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$ of point in a feature space to be classified in the frame of discernment $\Omega = \{w_0, w_1, \ldots, w_h\}$, where w_0 is explicitly included in the frame to represent a potential outlier class[1].

The center of each specific class can be obtained in many ways, like using the given data pdf model, the average of training data, the centers produced by an unsupervised clustering (estimation) method, etc. For example, FCM [1] is the most well-known unsupervised classification method working with the probability framework. In this work and in our experiments, we use the centers acquired by FCM. It is assumed that $C = \{\mathbf{c}_1, \ldots, \mathbf{c}_h\}$ are given and correspond to the centers of the specific classes $\Omega = \{w_1, \ldots, w_h\}$. The contribution of credal classification mainly lies in the introduction of meta-class, which is used to model imprecision of the class of the object.

In ECM [9], each class also corresponds to one class center, and the center of meta-class is the simple mean value of the involved specific classes' centers. The mass of belief on each class (specific class or meta-class) is proportional to only the distance between the object and the corresponding class center, and the distances to the centers of the involved specific classes are not taken into account in the determination of mass on the meta-class. The arithmetic mean value of the specific classes' centers generally cannot take the same distance to each center of the associated specific classes in many cases. Then, this center of meta-class seems distinguishable for its involved specific classes according to the distance measure. We argue that the specific classes included in a meta-class should be undistinguishable for the objects in this meta-class. Therefore, it is not a very effective way to use this mean value of vector to characterize the corresponding meta-class.

In this work, we propose a new center of meta-class to capture the imprecision of classes. If one object is closer to a center of meta-class (e.g. $w_i \cup \ldots \cup w_j$), these specific classes (e.g. $w_i, \ldots, w_j$) involved in the meta-class will be more undistinguishable for the object. So all the involved specific classes characterized by the corresponding centers should be undistinguishable for the center of this meta-class. Thus, this center of meta-class must keep the same distances to all the centers of the involved specific classes.

Let us consider a meta-class $U = w_i \cup \ldots \cup w_k$, and the center $\mathbf{c}_U$ of the meta class U is chosen at the same distance to all the centers of the specific classes included in U. Therefore

[1] It is worth noting that there is no class center corresponding to the outlier class w_0. The meta-classes involving w_0 do not enter in CCR because w_0 plays the role of the default (closure) class which will contain all data point that cannot be reasonably associated within $2^{\Theta \setminus \{w_0\}}$.

the following $\frac{1}{2}|U| \times (|U|-1)$ of conditions[2] must be satisfied

$$||\mathbf{c}_U - \mathbf{c}_i|| = ||\mathbf{c}_U - \mathbf{c}_j||, \quad \forall\ w_i, w_j \in U, i \neq j \qquad (1)$$

It is worth noting that $||.||$ refers to the Euclidean distance in this paper. Since Eq. (1) represents a set of $|U|-1$ independent constraints, there will be only one solution of $\mathbf{c}_U$ when the dimension of the vector $\mathbf{c}_U$ (i.e. the number of the attributes of data) is exactly $|U|-1$. If the dimension of $\mathbf{c}_U$ is bigger than $|U|-1$, there are many possible solutions for $\mathbf{c}_U$, and we will select the one which is the closest to all the centers of the specific classes included in U, i.e. such as $\arg[\min_{\mathbf{c}_U} \sum_{w_j \in U} (d(\mathbf{c}_U, \mathbf{c}_j))]$, since the objects in meta-class should be simultaneously close to all the involved specific classes. If the dimension of $\mathbf{c}_U$ is smaller than $|U|-1$, it will become an optimization problem to seek the solution of $\mathbf{c}_U$, which should be satisfied with the constraints as much as possible. It can be solved using any classical nonlinear optimization method. In this work, we seek the solution using the classical nonlinear least squares estimate method.

C. Construction of bba's

Let us consider one object $\mathbf{x}_s \in X, s = 1, \ldots, n$ to be classified under belief functions framework. If $\mathbf{x}_s$ is closer to a specific class center (e.g. $\mathbf{c}_i$), it indicates that $\mathbf{x}_s$ more probably belongs to the class w_i as done in the classical way. So the initial mass of $\mathbf{x}_s$ on a singleton class should be a monotone decreasing function (denoted by $f_1(\cdot)$) of the distance between the object and the corresponding class center. Then, one gets:

$$\tilde{m}(w_i) = f_1(d(\mathbf{x}_s, \mathbf{c}_i)), \forall i = 1, \ldots, h \qquad (2)$$

If one object is committed to a meta-class (e.g. $w_i \cup \ldots \cup w_j$), it means that this object must be simultaneously quite close to these specific classes (e.g. $w_i, \ldots, w_j$). Therefore, the distances between the object and each included specific class center should be taken into account. Besides that, these included specific classes should be quite undistinguishable for the object. Otherwise, this object will be assigned into a specific class rather than meta-class. So this object should be also very close to the center of the meta-class for which these specific classes are totally undistinguishable. As a result, we propose that the mass of belief for the object on a meta-class should be determined by both the distances to the centers (e.g. $\mathbf{c}_i, \ldots, \mathbf{c}_j$) of the specific classes involved in the meta-class and the distance to the center of this meta-class. In the determination of the mass of belief on a meta-class U for $\mathbf{x}_s$, the smaller distances between $\mathbf{x}_s$ and centers of the specific classes included in U means that $\mathbf{x}_s$ is more likely to belong to this set of classes (i.e. U), and the partial imprecision degree of the class of $\mathbf{x}_s$ is higher when $\mathbf{x}_s$ is closer to the center of U (i.e. $\mathbf{c}_U$). So the mass of belief for $\mathbf{x}_s$ on the meta-class U should be a function denoted by $f_2(\cdot)$ of both the distance to each center of the specific class involved in U and distance to center of meta-class $\mathbf{c}_U$.

[2] $|X|$ denotes the cardinality of the element X (i.e. the number of the singleton elements in X, for example $|A \cup B| = 2$).

$$\tilde{m}(U) = f_2(\Delta(\mathbf{x}_s, \mathbf{c}_U)) \qquad (3)$$

where

$$\Delta(\mathbf{x}_s, \mathbf{c}_U) \triangleq \frac{1}{|U|+\gamma} \cdot [\sum_{w_i \in U} d^\beta(\mathbf{x}_s, \mathbf{c}_i) + \gamma d^\beta(\mathbf{x}_s, \mathbf{c}_U)] \quad (4)$$

The tuning parameter β can be fixed to a small positive value (e.g. 1 or 2). The parameter γ is a tuning weighting factor of the distance between the object and the center of meta-class. The guidelines for tuning these parameters will be presented Sec. IV). $\Delta(\mathbf{x}_s, \mathbf{c}_U)$ represents the average value of the distances between $\mathbf{x}_s$ and all the centers of the involved specific classes, and the distance between $\mathbf{x}_s$ and $\mathbf{c}_U$. The function $f_2(.)$ should be decreasing with the increasing of $\Delta(\mathbf{x}_s, \mathbf{c}_U)$.

Both the functions $f_1(.)$ and $f_2(.)$ are monotone decreasing function, and they can be exactly determined according to the actual application under concern. In this work, we use the exponential decreasing function which is commonly used in classification problem [7]. The unnormalized mass of belief on specific class and meta-class are given by:

$$\tilde{m}(w_i) = e^{-\lambda d^\beta(\mathbf{x}_s, \mathbf{c}_i)} \qquad (5)$$

$$\tilde{m}(U) = e^{-\lambda \Delta(\mathbf{x}_s, \mathbf{c}_U)} \qquad (6)$$

λ is a parameter and it can be determined by the average distance between each pair of centers of the specific classes.

$$\lambda = \frac{1}{\frac{1}{2}|U|(|U|-1)} \cdot \sum_{i=1}^{n} \sum_{j>i} d^\beta(\mathbf{c}_i, \mathbf{c}_j) \qquad (7)$$

The outlier class denoted by w_0 is controlled by a given outlier threshold δ. If $\mathbf{x}_s$ is far from all the classes (i.e. distances to all classes are bigger than δ), it is considered that we can get little information about the the class of $\mathbf{x}_s$, and $\mathbf{x}_s$ is most possible the noisy (outlier) data. In practice, the outlier threshold δ can be decided according to the largest distance between the centers of specific classes, and it is defined by:

$$\delta = \eta \times \arg[\max_{i,j}(d^\beta(\mathbf{c}_i, \mathbf{c}_j))] \qquad (8)$$

where η is a positive tuning parameter. The unnormalized mass of belief of the outlier class for $\mathbf{x}_s$ is determined by:

$$\tilde{m}(w_0) = e^{-\lambda\delta} \qquad (9)$$

Because one usually works with normalized bba's in belief function theory, the unnormalized mass of belief $\tilde{m}(.)$ is finally normalized as follows $\forall A \subseteq \Omega$:

$$m(A) = \frac{\tilde{m}(A)}{\sum_{B \subseteq \Omega} \tilde{m}(B)} \qquad (10)$$

This normalized mass function (bba) $m(.)$ is then used for the credal classification of the object $\mathbf{x}_s$.

III. CCR TAKING INTO ACCOUNT THE DATA DISPERSION

The dispersion degree of different data classes can be quite different in some real applications, and this should be taken into account to improve the classification performance when it is available. For instance, in the supervised data classification problem, the training data samples are available to estimate the variance of each data class, which can represent the dispersion degree. It is assumed that the variance matrix of each class of data is Σ, which is considered as a diagonal matrix without lack of generality, and it is defined by:

$$\Sigma = \begin{pmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & \sigma_h^2 \end{pmatrix} \tag{11}$$

A. The metric center of meta-class

In the determination of metric center of meta-class U, a small standard deviation of a data class (e.g. $w_i \in U$) indicates that the dispersion of w_i is small. In this case, the meta-class center $\mathbf{c}_U$ should be closer to the class w_i. For this reason, the distance $d(\mathbf{c}_U, \mathbf{c}_i)$ between $\mathbf{c}_U$ and the specific class center $\mathbf{c}_i$ is chosen proportionally to the standard deviation of the class w_i. A bigger standard deviation σ_i leads to a bigger distance $d(\mathbf{c}_U, \mathbf{c}_i)$.

Let us consider a meta-class U with the center of $\mathbf{c}_U$. One should find a solution of $\mathbf{c}_U$ that satisfies the set of equations given in Eq. (12) as best as possible.

$$\frac{d(\mathbf{c}_U, \mathbf{c}_i)}{d(\mathbf{c}_U, \mathbf{c}_j)} = \frac{\sigma_i}{\sigma_j}, \forall w_i, w_j \in U, i \neq j \tag{12}$$

There are $\frac{1}{2}|U| \times (|U| - 1)$ equations with different pairs of $\mathbf{c}_i$ and $\mathbf{c}_j$ as Eq. (12), and $\mathbf{c}_U$ can be solved similarly as for Eq. (1).

B. Determination of bba's with variance of data class

The unnormalized mass of belief committed to the specific class as Eq. (5) can be modified taking into account the variance of each class by:

$$\tilde{m}(w_i) = e^{-\frac{d^\beta(\mathbf{x}_s, \mathbf{c}_i)}{\sigma_i^\beta}}, \forall w_i \in \Omega \tag{13}$$

Similarly, the unnormalized mass of belief assigned to the meta-class U is given by:

$$\tilde{m}(U) = e^{-\bar{\Delta}(\mathbf{x}_s, \mathbf{c}_U)} \tag{14}$$

with

$$\bar{\Delta}(\mathbf{x}_s, \mathbf{c}_U) \triangleq \frac{1}{|U| + \gamma} \cdot [\sum_{w_i \in U} \frac{d^\beta(\mathbf{x}_s, \mathbf{c}_i)}{\sigma_i^\beta} + \gamma \frac{d^\beta(\mathbf{x}_s, \mathbf{c}_U)}{\sum_{w_i \in U} \sigma_i^\beta / |U|}] \tag{15}$$

In the determination of mass on the outlier class, the given threshold ξ can be tuned according to the maximal variance of all the classes.

$$\tilde{m}(w_0) = e^{-\xi} \tag{16}$$

with

$$\xi = \rho \times \frac{\max([\sigma_1^\beta, \ldots, \sigma_n^\beta])}{\sum_{i=1}^{n} \sigma_i^\beta / n} \tag{17}$$

where ρ plays the similar role as η in Eq. (8).

This unnormalized mass function $\tilde{m}(.)$ is normalized according to Eq. (10).

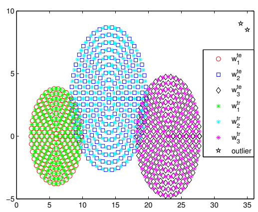
IV. SIMULATION RESULTS

In this section we present two experiments based on artificial data and real data to show the performances of CCR. The experiments # 1 used artificial (simulated) data for a 3-class problem is to qualitatively illustrate the use of credal classification in CCR. The experiment # 2 shows the performance of CCR based on two real-data sets with respect to several other methods.

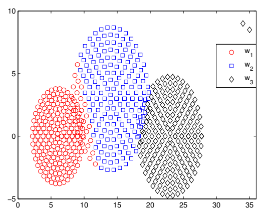
Guidelines for tuning the parameters in CCR: The parameters β, γ, ρ, and η must be tuned for applying CCR. The parameter β can take a small positive integer value (e.g. 1, 2) and has a very little influence on the performance of CCR. In the simulations, we have taken $\beta = 2$. The parameter γ is used to control the number of objects in the meta-classes. The bigger value of γ will produce more objects committed to the meta-classes. Generally, one can take $\gamma \in (0, 5)$, but if needed γ can be tuned according to the imprecision degree (i.e. the rate of the objects in meta-class) one expects. The parameter η is associated with the outlier threshold when the data dispersion is not taken into account. The bigger η will cause smaller number of outliers, and we generally recommend to take $\eta \in [0.5, 2]$. ρ plays the similar role as η when the variances of data classes are available, and $\rho \in [2, 5]$ can be considered as default ranges. Both η and ρ should be determined by the outlier rate that one wants to tolerate in the actual application. The exact value of these parameters can be optimized in the training data space (if available), and the optimized value should correspond to the good classification results. We select that $\eta = 2$ in Experiment # 2 for case 1, and $\rho = 3$ in Experiment #1 and Experiment # 2 for case 2.

A. Experiment #1: A 3-classes problem with artificial data

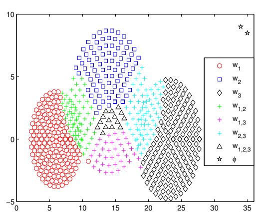
We consider a particular 3-class data set in a circular shape as shown in 1-a in this experiment. This data set consists of 615 training data points and 617 test data points including 2 noisy points (outliers). The radii of the circles for w_1, w_2 and w_3 are $r_1 = 4, r_2 = 6, r_3 = 5$. CCR is applied for the classification of this particular data set and it is compared with the classical unsupervised FCM and ECM classification methods, and also with the supervised K-nearest neighbor (K-NN) [13]. FCM and ECM are directly applied for the classification of the test data. A particular value of $K = 11$ is selected in K-NN, since it provides a good result. In CCR, the training data set is used to get the center and the variance of each class. Then, the test data is classified using CCR. Different values for γ (0.5 and 1) used in CCR have been tested to evaluate their impacts on the final results. The classification results obtained by different methods are shown in Fig. 1-(b–f). We denote $w^{te} \triangleq w^{test}$, $w^{tr} \triangleq w^{training}$ and $w_{i,\cdots,k} \triangleq w_i \cup \cdots \cup w_k$ for convenience. The running time t is indicated in the title of each subfigure.



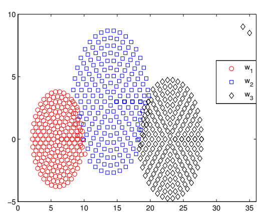
(a). Original data to classify.



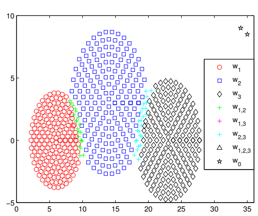
(b). Classification result by FCM (t=0.5616s).



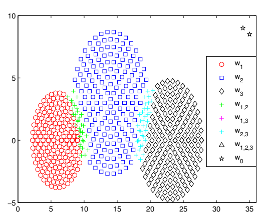
(c). Classification result by ECM with $\alpha = 1$ (t=7.4412s).



(d).Classification result by K-NN with $K = 11$ (t=2.5740s).



(e). Classification result by CCR with $\gamma = 0.5$ (t=0.6396s).



(f). Classification result by CCR with $\gamma = 1$ (t=0.6552s).

Figure 1. Classification results by different methods for a 3-class problem.

On Fig. 1-(a), we see that the class w_2 partly overlaps with w_1 and w_3, and the points belonging to the overlapped zones are really difficult to classify correctly due to their ambiguity. As shown on Fig. 1-(b), FCM produces only three singleton clusters w_1, w_2 and w_3 based on the probability framework, and most points in the overlapped zone of w_1 (resp. w_3) and w_2 are committed to w_1 (resp. w_3) which can cause many misclassification errors. Moreover, the outliers can not be detected by FCM. ECM provides the credal partitions in belief functions framework. We can see that w_1 and w_3 are not close and they are totally separate, but there are still many points originally from w_2 that are wrongly committed to the meta-cluster $w_1 \cup w_3$ labeled by blue plus symbol in Fig. 1-(c),(d). More dramatically, many points from w_2 labeled by green hexagon are even associated with the total ignorant cluster $w_1 \cup w_2 \cup w_3$. These unreasonable results show the wrong behavior of ECM. The supervised classifiers K-NN produce the similar results as FCM. Most of objects in the overlapped zone of different classes are simply classified into a particular class (i.e. w_1, w_3), and the outliers are wrongly committed to the class w_3. This solution is not very effective, and it will yield many misclassification errors. CCR produces more reasonable credal classification results in comparison with other methods. The points in the middle of w_1 and w_2, w_2 and w_3 are respectively committed to $w_1 \cup w_2$ and $w_2 \cup w_3$ as shown in Fig. 1-(e)–(f) since these points are really hard to be correctly classified into a particular class and these specific classes seem undistinguishable for these objects. The outliers are well detected by CCR. Moreover, CCR requires much smaller computation burden than ECM according the running time t indicated in the title of each subfigure. One can see that he bigger γ value in CCR leads to more data points in meta-class as shown on Fig. 1-(e)-(f). So the γ value should be tuned to obtain a good compromise between the imprecision degree and the misclassification rate.

B. Experiment #2: Real data sets

In this experiment, we use two well-known real data sets (i.e. iris data set and wine data set) from UCI Repository (http://archive.ics.uci.edu/ml/) to test the performance of CCR with respect to other methods. The Iris data set contains 3 classes with 150 samples, and each sample has 4 attributes. Wine data set consists of 178 samples with 13 attributes for each sample. In this experiment, two analysis are conducted:

- Case 1: The two data sets are applied for the unsupervised data classification using FCM, ECM and CCR. In CCR, the class centers obtained by FCM are adopted. Different values of α in ECM and γ in CCR are selected to show their influence on the results.
- Case 2: The k-fold cross validation is performed on the two data sets by ANN, K-NN and CCR, and we use the simplest 2-fold cross validation here. In K-NN, we give the average error rate with the number of K ranking from 5 to 15, and we use the feed-forward back propagation network for ANN, with $epochs = 1000$ and $goal = 0.001$.

In this real data sets, the belief on the ignorance (outlier) class in CCR are proportionally redistributed to the other available focal elements. The classification results of the two data sets by different methods are respectively shown in Tables I–II (where ”NA” means ”Not applicable”). For one object originated from w_i, if it is classified into A with $w_i \cap A = \emptyset$, it will be considered as an error of classification. If $w_i \cap A \neq \emptyset$ and $A \neq w_i$, it will be considered as an imprecise classification. The error rate R_e is calculated by $R_e = N_e/T$, where N_e is number of objects wrongly classified, and T is the total number of the objects tested. The imprecision rate R_{I_j} is calculated by $R_{I_j} = N_{I_j}/T$, where N_{I_j} is number of objects committed to the meta-classes with the cardinality value j.

When there is no training data (as considered in case 1), FCM produces the biggest error rate because of the limitations of the probability framework. The number of errors generated by ECM is a bit less than with FCM, but it causes a high imprecision degree of the results, which is not an efficient solution. CCR provides a good compromise between the error rate and the imprecision rate. For the iris data set, CCR

Table I
CLASSIFICATION RESULTS OF IRIS DATA WITH DIFFERENT METHODS (IN%)

	method	parameter	R_e	R_{I_2}	R_{I_3}
case 1	FCM		10.67	NA	NA
	ECM	α=2.0	8.00	4.67	0
	ECM	α=1.5	10.0	8.67	0.67
	ECM	α=1.0	10.0	15.33	6.00
	CCR	γ=1.0	7.33	6.00	0
	CCR	γ=1.5	4.67	10.0	0
	CCR	γ=2.0	3.33	12.0	0
case 2	ANN		4.00	NA	NA
	K-NN		3.87	NA	NA
	CCR	$\gamma = 0.5$	1.33	6.67	0
	CCR	$\gamma = 1.0$	1.33	12.00	0
	CCR	$\gamma = 2.0$	1.33	13.33	0

Table II
CLASSIFICATION RESULTS OF WINE DATA WITH DIFFERENT METHODS(IN%)

	method	parameter	R_e	R_{I_2}	R_{I_3}
case 1	FCM		31.46	NA	NA
	ECM	α=2.0	16.29	30.34	3.37
	ECM	α=4.0	23.03	20.79	0.56
	ECM	α=5.0	26.97	13.48	0.56
	CCR	γ=1.0	26.97	8.99	0
	CCR	γ=2.0	23.60	15.17	0
	CCR	γ=3.0	21.35	21.91	0
case 2	ANN		33.71	NA	NA
	K-NN		27.73	NA	NA
	CCR	$\gamma = 0.5$	24.72	5.06	0
	CCR	$\gamma = 1.0$	22.47	10.11	0
	CCR	$\gamma = 2.0$	22.47	12.92	0

produces smaller error rate than ECM and FCM, and the imprecision degree of CCR is generally smaller than with ECM. For the wine data set, when ECM and CCR produce similar error rates with the given parameter α and γ, the imprecision degree of CCR remains smaller than with ECM. This first analysis (case 1) shows that CCR provides better performances than ECM and FCM.

In case 2, the training data set is considered, and the classical ANN and K-NN supervised classification methods are included in our performances comparison. One can see that the error rate obtained by CCR is smaller that that by ANN and K-NN methods, since these objects hard to correctly classify have been committed to the meta-class by CCR. Besides that, the computation burden of CCR is much smaller than ECM, ANN and K-NN.

This experiment also shows that the bigger γ value leads to bigger number of objects assigned to meta-class. So γ should be tuned according to the imprecision degree one expects. The results shown in Table II indicate that some objects are committed to meta-classes by CCR. This means that the used attributes are not sufficient for the specific classification of these objects (data points). CCR alerts us that we should treat these objects in meta-classes more cautiously.

V. CONCLUSION

A simple and effective credal classification rule (CCR) has been proposed based on belief functions for the classification of uncertain data. CCR strengths the robustness of results by reducing the misclassification errors thanks to the introduction of meta-classes corresponding to the disjunction of several specific classes. CCR is also able to detect the outlier points. In CCR, an object quite close to a specific class center is committed to this specific class. If the object is simultaneously close to several specific classes (and these specific classes seem undistinguishable for this object), then the object should be cautiously assigned to a proper meta-class. The objects too far from the others with respect to a given threshold will be reasonably considered as outliers. The output of CCR is not necessarily used for a final classification purpose, but it can be used efficiently to alert the classification system designer that some other complementary information sources are required to remove ambiguity of some particular data points. The potential and interest of CCR have been shown using both the artificial and real data sets.

Acknowledgements

This work has been partially supported by National Natural Science Foundation of China (No. 61075029, 61135001).

REFERENCES

[1] J. C. Bezdek, *Pattern recognition with fuzzy objective function algorithms*, Plenum Press, New-York, 1981.
[2] C.M. Bishop, *Neural networks for pattern recognition*, Oxford: Oxford University Press, 1995.
[3] A.-L. Jousselme, C. Liu, D. Grenier, E. Bossé, *Measuring ambiguity in the evidence theory*, IEEE Trans on SMC-Part A, Vol. 36(5), pp. 890–903, Sept. 2006.
[4] G. Shafer, *A mathematical theory of evidence*, Princeton Univ. Press, 1976.
[5] F. Smarandache, J. Dezert (Editors), *Advances and applications of DSmT for information fusion*, American Research Press, Rehoboth, Vol. 1-3, 2004-2009. http://fs.gallup.unm.edu/DSmT.htm
[6] P. Smets, *The combination of evidence in the transferable belief model*, IEEE Trans. Pattern Anal. Mach. Intell., Vol. 12(5), pp. 447–458,1990.
[7] T. Denœux, *A k-nearest neighbor classification rule based on Dempster-Shafer Theory*, IEEE Trans. on Systems, Man and Cybernetics, Vol. 25(05), pp. 804–813,1995.
[8] Z.-g. Liu, J. Dezert, G. Mercier, Q. Pan, *Belief C-Means: An extension of fuzzy c-means algorithm in belief functions framework*, Pattern Recognition Letters, Vol. 33(3), pp. 291–300, 2012.
[9] M.-H. Masson, T. Denœux, *ECM: An evidential version of the fuzzy c-means algorithm*, Pattern Recognition, Vol. 41(4), pp. 1384–1397, 2008.
[10] T. Denœux, M.-H. Masson, *EVCLUS: EVidential CLUStering of proximity data*, IEEE Trans. on Systems, Man and Cybernetics Part B, Vol. 34(1), pp. 95–109, 2004.
[11] T. Denœux, P.Smets, *Classification using belief functions: relationship between case-based and model-based approaches*, IEEE Trans. on Systems, Man and Cybernetics, Part B: Vol. 36(6), pp. 1395-1406, 2006.
[12] T. Denoeux, *Maximum likelihood estimation from uncertain data in the belief function framework*, IEEE Transactions on Knowledge and Data Engineering, Vol. 25(1), pp. 119–130, 2013.
[13] B.W. Silveman, M.C. Jones, *E. Fix and J. L. Hodges (1951): An important contribution to nonparametric discriminant analysis and density estimation*, Int. Statistical Review, Vol. 57(3), pp. 233–247, 1989.