



Credal classification rule for uncertain data based on belief functions

Zhun-Ga Liu, Quan Pan, Jean Dezert, Grégoire Mercier

► To cite this version:

Zhun-Ga Liu, Quan Pan, Jean Dezert, Grégoire Mercier. Credal classification rule for uncertain data based on belief functions. Pattern Recognition, 2014, 47 (7), pp.2532-2541. 10.1016/j.patcog.2014.01.011 . hal-01060091

HAL Id: hal-01060091

<https://hal.science/hal-01060091>

Submitted on 14 Jun 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

Credal classification rule for uncertain data based on belief functions

Zhun-ga Liu^{a,b,*}, Quan Pan^a, Jean Dezert^c, Grégoire Mercier^b

^a School of Automation, Northwestern Polytechnical University, Youyi xilu 127, Xi'an, Shaanxi, 710072, PR

China ^b Telecom Bretagne, CNRS UMR 6285 Lab-STICC/CID, Brest, France

^c ONERA – The French Aerospace Lab, F-91761 Palaiseau, France

In this paper we present a new credal classification rule (CCR) based on belief functions to deal with the uncertain data. CCR allows the objects to belong (with different masses of belief) not only to the specific classes, but also to the sets of classes called meta-classes which correspond to the disjunction of several specific classes. Each specific class is characterized by a class center (i.e. prototype), and consists of all the objects that are sufficiently close to the center. The belief of the assignment of a given object to classify with a specific class is determined from the Mahalanobis distance between the object and the center of the corresponding class. The meta-classes are used to capture the imprecision in the classification of the objects when they are difficult to correctly classify because of the poor quality of available attributes. The selection of meta-classes depends on the application and the context, and a measure of the degree of indistinguishability between classes is introduced. In this new CCR approach, the objects assigned to a meta-class should be close to the center of this meta-class having similar distances to all the involved specific classes' centers, and the objects too far from the others will be considered as outliers (noise). CCR provides robust credal classification results with a relatively low computational burden. Several experiments using both artificial and real data sets are presented at the end of this paper to evaluate and compare the performances of this CCR method with respect to other classification methods.

1. Introduction

The classical methods of classification have been developed at first in the probability theory framework. These methods compute the probability assignments of the objects in different specific classes. The final assignment (classification) of an object is determined by the class committed with the highest probability value. In the classification of uncertain data, the different classes can partly overlap, and the objects in the overlapped zones are really hard to be correctly classified into a particular class due to the insufficient attributes information. Probability theory framework are not well adapted to characterize such uncertainty and imprecision [1–3].

The belief functions (BF) [4–8] introduced in Dempster-Shafer theory (DST) have been widely used to model the uncertain and imprecise information for data clustering [9–11], data classification [12–17], image processing [18,19], and for information fusion [20–22]. A new concept, called *credal partition*, based on belief functions for the unsupervised data clustering has been introduced by Denœux and Masson in [10]. The credal partitioning allows the objects to belong to the specific classes, and to the sets of classes with different belief mass assignments. This provides a deeper insight in the data. An Evidential

CLUstering (EVCLUS) [10] algorithm working with credal partition has been developed for relational data. An Evidential C-Means (ECM) [9] clustering method inspired from the Fuzzy C-Means (FCM) [23], and a Noise-Clustering algorithm [24] have also been proposed for the credal partition of object data. However, ECM can produce very unreasonable results when the different classes' centers are sufficiently close. This serious drawback has been clearly shown and discussed in [11]. In our previous related works, we have developed a method called belief C-means (BCM) [11] to overcome the limitation of ECM by adopting another interpretation of the meta-class. An evidential EM algorithm [25] has been recently proposed for the parameter estimation in statistical models when the uncertainty on the data can be modeled by belief functions. Some supervised data classification methods [15] have been also developed based on DST. The model-based classifiers [15] have been proposed by Denœux and Smets based on Smets' transfer belief model (TBM) [6–8]. An evidential version of K -nearest neighbors rule (EK-NN) is proposed in [13], and a neural network classifier based on DST is presented in [26]. All these evidential classifiers consider only as possible assignment solution the specific classes, and one extra class (i.e. the ignorant class) which is defined by the disjunction of all the specific classes. In these supervised methods, the meta-classes¹ (i.e. the partially ignorant classes) are not considered as useful solutions of the credal classification.

* Corresponding author. Tel.: +86 2988431366.

E-mail addresses: liuzhunga@gmail.com (Z.-g. Liu), jean.dezert@onera.fr (J. Dezert), Gregoire.Mercier@telecom-bretagne.eu (G. Mercier).

¹ Defined by the disjunction of several specific classes.

In our opinion, the meta-classes play an important role to characterize the imprecision of classification of the objects. The objects hard to classify should be reasonably committed to the meta-class, which can well reflect the imprecision (ambiguity) degree of the classification, and reduce the misclassification errors as well. In our very recent work, a belief K -nearest neighbor (BK-NN) classifier [14] working with credal classification has been developed to deal with uncertain data by considering all possible meta-classes in the process. Such method however requests a high computational burden which is usually the main drawback of all K -NN alike methods [27]. The purpose of this paper is to propose a new straightforward and more simple mathematical solution, called Credal Classification Rule (CCR), for computing the basic belief assignments of uncertain data for their credal classification.

The interest of credal classification mainly resides in its ability to commit objects to the meta-classes rather than to the specific classes when the information is insufficient for making it correctly. By doing so, we preserve the robustness of the result and we 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, like in target classification and tracking, it is very crucial to maintain such robustness than to provide immediately (with high risk of error) a precise classification. The credal classification can be very helpful to manage external (possibly costly) complementary resources in order to reduce some particular ambiguities. Our approach is very helpful for requesting (or not) a complementary information sources (if possible and available) in order to get more precise reliable classification results, and to reduce dramatic errors in the final decision-making process.

In this new CCR approach, each specific class is characterized by the corresponding class center (i.e. prototype) computed from the training data. The center of a meta-class is calculated based on the centers of specific classes included in the meta-class. In the multi-class classification problem, there are usually only few (not all) classes that partly overlap, and most classes that are in fact far from each other can be well separated. The meta-class defined by the union of the classes that are far from each other are not useful in such applications. In order to reduce the computational complexity, we just need to select the useful meta-classes according to the context of the application under concern. The belief mass assignment of the object to classify with each specific class is determined based on the Mahalanobis distance between the object and the corresponding specific class center. Intuitively, the object committed to a specific class should be very close its center. If the object to classify is assigned to a meta-class, it means that the true class of the object is among the specific classes included in the meta-class but we do not know which one precisely. The ratio of the maximum distance of the object to the involved specific classes' centers, over the minimum distance, is introduced to measure the degree of distinguishability of these classes. Thus, the belief mass of a meta-class is determined from the distance between the object and the center of meta-class and its corresponding ratio value. An object will be committed to a meta-class with a high belief mass as soon as it is located at (almost) the same distances of several specific classes centers. Because in that case, it means that the object is very difficult to be correctly classified into a specific class. CCR provides credal classification results with low computational burden due to the simple working principle.

After a brief presentation of belief functions in Section 2, we state in Section 3 the principles of CCR and the mathematical computation of bba's for the credal classification. In Section 4, we present some classification results based on artificial and real data sets, and we compare the performances of the CCR with respect to well-known classification methods. Conclusions are given in Section 5.

2. Basics of belief functions theory

The belief functions have been introduced by Shafer in 1976 in his Mathematical Theory of Evidence known also as Dempster-Shafer Theory (DST) [4–8]. Let us consider a finite discrete set $\Theta = \{\theta_1, \theta_2, \dots, \theta_h\}$ of $h > 1$ mutually exclusive and exhaustive hypotheses θ_i , $i = 1, 2, \dots, h$. This set Θ is called the *frame of discernment* of the problem under consideration. The power-set of Θ , denoted 2^Θ , includes all the subsets of Θ . It is defined by

$$2^\Theta = \{A | A \subseteq \Theta\} \quad (1)$$

For example, if $\Theta = \{\theta_1, \theta_2, \theta_3\}$, then $2^\Theta = \{\emptyset, \theta_1, \theta_2, \theta_3, \theta_1 \cup \theta_2, \theta_1 \cup \theta_3, \theta_2 \cup \theta_3, \theta_1 \cup \theta_2 \cup \theta_3\}$.

In the frame of discernment Θ , each element (e.g. $\theta_i \in \Theta$) represents one single hypothesis, and it characterizes one class in this work. The union $\theta_i \cup \theta_j \equiv \{\theta_i, \theta_j\}$ of two elements² θ_i and θ_j 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 , and θ_i and θ_j are undistinguishable”.

A basic belief assignment (bba) is a function $m(\cdot)$ from 2^Θ to $[0, 1]$ satisfying

$$\sum_{A \in 2^\Theta} m(A) = 1 \quad (2)$$

The subsets A of Θ such that $m(A) > 0$ are called the *focal elements* of $m(\cdot)$. The *credal partition* [9,10] is defined as n -tuple $M = (\mathbf{m}_1, \dots, \mathbf{m}_n)$, where \mathbf{m}_i is the basic belief assignment of the object $\mathbf{x}_i \in X$, $i = 1, \dots, n$ 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.

From any bba $m(\cdot)$, the belief function $Bel(\cdot)$ and the plausibility function $Pl(\cdot)$ are defined for $\forall X \in 2^\Theta$ as

$$\begin{cases} Bel(X) = \sum_{Y \in 2^\Theta | Y \subseteq X} m(Y) \\ Pl(X) = \sum_{Y \in 2^\Theta | X \cap Y \neq \emptyset} m(Y) \end{cases} \quad (3)$$

$Bel(X)$ represents the whole mass of belief that comes from all subsets of Θ included in X . It is interpreted as the lower bound of the probability of X , i.e. $P_{\min}(X)$. $Bel(\cdot)$ is a sub-additive measure since $\sum_{\theta_i \in \Theta} Bel(\theta_i) \leq 1$. $Pl(X)$ represents the whole mass of belief that comes from all subsets of Θ compatible with X (i.e. those intersecting X). $Pl(X)$ is interpreted as the upper bound of the probability of X , i.e. $P_{\max}(X)$.

The Pignistic probability (or betting probability) transformation $BetP(\cdot)$ introduced by Smets [6,7] is commonly used to transform any bba $m(\cdot)$ into a probability measure for the decision-making support based on the maximum of $BetP(\cdot)$ value. Mathematically, $BetP(A)$ is defined $\forall A \in 2^\Theta \setminus \{\emptyset\}$ by

$$BetP(A) = \sum_{B \in 2^\Theta, A \subseteq B} \frac{|A \cap B|}{|B|} m(B) \quad (4)$$

where $|X|$ is the cardinality of the element X (i.e. the number of the singleton elements in X , for example if $X = \theta_i \cup \theta_j$ then $|X| = 2$).

In DST [4], the combination of distinct bba's is done by Dempster's rule of combination. This paper only focuses on the construction of bba $m(\cdot)$ in the credal classification context and does not concern the combination of bba's.

² Since there is one-to-one mapping between propositions and sets [4], the union set operator is equivalent to the disjunction operator of propositions. Hence, $\theta_1 \cup \theta_2 \cup \dots \cup \theta_h \equiv \Theta$.

3. Credal classification rule (CCR)

In this section we present in detail the Credal Classification Rule (CCR) for classifying uncertain data. CCR provides a simple and an efficient way to compute the belief mass of the assignment of the object with the specific classes, with the meta-classes (which characterize the partial imprecision of class), and with the outlier class (i.e. the full ignorant class). The CCR consists of two main steps: (1) the determination of centers of the specific and meta classes, and (2) the construction of the bba's based on the distances between the object and each class center.

3.1. Determination of the centers of classes

Let us consider a given set of data $X = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n\}$, where the vectors $\mathbf{x}_i (i=1, \dots, n)$ have to be classified over a frame of discernment $\Omega = \{w_0, w_1, \dots, w_h\}$ using the training data set $Y = \{\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_g\}$. The element w_0 is explicitly included in the frame Ω to represent the unknown extra class for the closure of the frame.

The center of each specific class of Ω can be obtained in many ways.³ For instance, one can use a given data pdf model, or the average of training data, or the centers produced by an unsupervised clustering (estimation) method (e.g. FCM, EM, etc). In this work, the center of each specific class is simply defined by the mean value of the training data $Y = \{\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_g\}$ in the corresponding class. It is assumed that $C = \{\mathbf{c}_1, \dots, \mathbf{c}_h\}$ are given, and correspond to the centers of the specific classes w_1, \dots, w_h . For $j = 1, 2, \dots, h$, the center \mathbf{c}_j is defined $\forall \mathbf{y}_i \in w_j$ by

$$\mathbf{c}_j = \frac{1}{S_j} \sum_{\mathbf{y}_i \in w_j} \mathbf{y}_i \quad (5)$$

where S_j is the number of training samples in the class w_j .

The interest of the credal classification is taking into account of the meta-classes that are used to model the imprecision of the class of the object to classify. In the ECM [9] method, each class is characterized by its class center, and the center of meta-class is the simple mean value of the involved specific classes' centers. The mass of belief committed to each class is proportional to the distance between the object and the corresponding class center. The bigger distance leads to the smaller mass of belief. Unfortunately, the determination of the bba proposed by the ECM approach is not very efficient because the mean value of vector cannot well represent the meta-class. Indeed, the arithmetic mean value of the specific classes' centers generally does not take the same distance to the each center of the associated specific class. Because of this, the centers of the involved specific classes become distinguishable of the center of the meta-class if one applies the ECM approach. This ECM behavior is not very good because the meta-class should fairly reflect the real impossibility to distinguish the involved specific classes for the object belonging to this meta-class.

To palliate the aforementioned major drawback of ECM, we propose a new method to determine the center of the meta-classes. Basically in our approach, an object to classify will be committed to a meta-class (e.g. $w_i \cup w_j \dots \cup w_k$), as soon as all the specific classes (e.g. w_i, w_j, \dots, w_k) become undistinguishable for this object according to the distance measures. Therefore, we argue that the center of a meta-class must be located at the same distances of all the centers of the specific classes included in the meta-class under consideration.

³ 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 points that cannot be reasonably associated within $2^{\Theta \setminus \{w_0\}}$.

For instance, let us consider the simplest meta-class (e.g. $U = w_i \cup w_j$) having a cardinality equals to two, e.g. $|U| = 2$. The meta-class center, denoted \mathbf{c}_U , should be at the same distance to all the specific classes' centers include in U , which are \mathbf{c}_i and \mathbf{c}_j . Therefore, the following condition must be satisfied:

$$d(\mathbf{c}_U, \mathbf{c}_i) = d(\mathbf{c}_U, \mathbf{c}_j) \quad (6)$$

Eq. (6) represents only one constraint, and it can produce 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 one. If the dimension of \mathbf{c}_U is bigger than one, there are many possible solutions for \mathbf{c}_U . Then, we will select the one which is closest to all the centers of the specific classes included in U , and given by $\arg[\min_{\mathbf{c}_U} \sum_{w_j \in U} (d(\mathbf{c}_U, \mathbf{c}_j))]$ because the meta-class center should be also simultaneously close to all the involved specific classes as much as possible.

It is worth noting that Mahalanobis distance (i.e. the normalized Euclidean distance) is used in this work to deal with the anisotropic data sets. This distance between two vectors \mathbf{c}_U and \mathbf{c}_i is given by

$$d_{Ui} \triangleq d(\mathbf{c}_U, \mathbf{c}_i) = \sqrt{\sum_{k=1}^N \frac{(\mathbf{c}_U(k) - \mathbf{c}_i(k))^2}{\delta_i(k)^2}} \quad (7)$$

where N is the number of dimensions (attributes/features) of \mathbf{c}_U and \mathbf{c}_i , and $\delta_i(k)$ is the standard deviation of the training data of class w_i in its k -th dimension.

The object committed to a meta-class (e.g. $U = w_i \cup w_j$) indicates that it must truly belong to one of the specific classes included in this meta-class, but these specific classes are not very distinguishable for this object. So the meta-class center should be closer to the centers of these involved specific classes than to other incompatible classes' centers.⁴ Therefore, the following condition must be satisfied:

$$\max_{w_i \in U} d_{Ui} < \min_{w_k \notin U} d_{Uk} \quad (8)$$

If the condition given by Eq. (8) is fulfilled, then one considers that the meta-class U must be kept as a potential solution of the classification, i.e. as a focal element of the bba. Otherwise, if the meta-class center is closer to the center of an incompatible specific class $w_k \notin U$ than to the center of a specific class $w_i \in U$, it indicates that the objects close around the center \mathbf{c}_U should belong more likely to the specific class $w_k \notin U$ rather than to $w_i \in U$. In such case, this meta-class U cannot be considered as effective⁵ for the classification solution, and the center \mathbf{c}_U should be eliminated to reduce the computational burden by reducing the number of focal elements of the bba.

Fig. 1 illustrates the selection of the meta-class. One considers a three class problem with $\Omega = \{w_1, w_2, w_3\}$ and the corresponding set of centers $\{\mathbf{c}_1, \mathbf{c}_2, \mathbf{c}_3\}$ as shown in Fig. 1. One sees that the class w_2 partly overlaps w_1 and w_3 , whereas w_1 and w_3 are well separated. The meta-class center $\mathbf{c}_{1,2}$ is more close to \mathbf{c}_1 and \mathbf{c}_2 than to \mathbf{c}_3 . So the meta-class⁶ $w_1 \cup w_2$ is considered acceptable and its center $\mathbf{c}_{1,2}$ should be kept for the determination of the mass of belief (see Section 3.2 for details). For a similar reason, the meta-class $w_2 \cup w_3$ with the center $\mathbf{c}_{2,3}$ is also acceptable. However, the center $\mathbf{c}_{1,3}$ is closer to the incompatible class' center \mathbf{c}_2 than to \mathbf{c}_1 and to \mathbf{c}_3 which indicates that the objects around $\mathbf{c}_{1,3}$ will more

⁴ The elements A and B are considered incompatible if $A \cap B = \emptyset$, and compatible if $A \cap B \neq \emptyset$.

⁵ Because it is very likely that the specific classes (e.g. w_i and w_j) in U are separated by the class w_k . The classes w_i and w_j in fact do not overlap, and so no object need to be assigned to the meta-class $U = w_i \cup w_j$.

⁶ Actually according to Fig. 1, the overlapped zone between w_1 and w_2 should better correspond to $w_1 \cap w_2$. Because in this work we do not allow an object to belong simultaneously to several distinct classes, the object in the overlapped zone is supposed to belong to one class only, but we cannot exactly determine precisely which class (w_1 or w_2). So the meta-class $w_1 \cup w_2$, which is more coherent with our interpretation here than $w_1 \cap w_2$, is used to represent the overlapped zone.

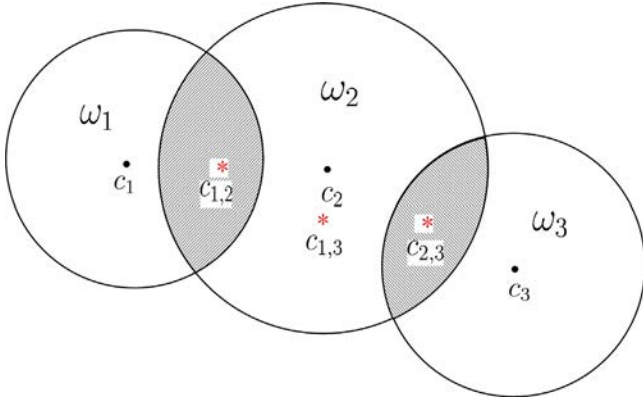


Fig. 1. Simple illustration of the meta-class selection.

likely belong to the class w_2 . Consequently, the meta-class $w_1 \cup w_3$ will not be taken into account in the credal classification, and its center $c_{1,3}$ is useless for the determination of the bba because one will take $m(w_1 \cup w_3) = 0$ in that case.

Let us consider the more general situation with the cardinality value of the meta-class bigger than two (i.e. $|U| \geq 3$). If the meta-class U is accepted⁷ in the credal classification, it indicates that all the specific classes included in U should be undistinguishable for the objects committed to this meta-class. So all the subsets (i.e. the sub-meta-classes) of U should be also acceptable before entering the calculation of meta-class center c_U . If one meta-class $A \subset U$ is considered unacceptable, it means that several specific classes in A can be distinguished by all the objects, and there is no necessity to preserve the meta-class A as a focal element of the bba. In that case, the meta-class U of course becomes unacceptable (useless), and we do not need to calculate its center. If all the subsets of U are acceptable, then one can go for the computation of the center c_U to determine $m(U) > 0$.

Because the center c_U must be located at the same Mahalanobis distance with respect to all centers of the specific classes included in U , the following conditions must hold:

$$d(c_U, c_i) = d(c_U, c_j), \quad \forall w_i, w_j \in U, \quad i \neq j. \quad (9)$$

Since one can obtain a set of $|U| - 1$ independent constraints from Eq. (9), there will be only one solution of c_U when the number of the available attributes of data is equal to $|U| - 1$. If the number of attributes is bigger than $|U| - 1$, there exist many solutions for c_U . If so, we will select the solution which is closest to all the centers of the specific classes included in U , and given by $\arg[\min_{c_U} \sum_{w_j \in U} (d(c_U, c_j))]$. If the dimension of c_U is smaller than $|U| - 1$, one has to solve an optimization problem to seek the solution for c_U that should be satisfied with all the constraints as much as possible, such as for $\forall w_i, w_j \in U, i \neq j$, $\arg[\min_{c_U} \sum_{w_i, w_j \in U} (d(c_U, c_i) - d(c_U, c_j))^2]$. This can be done using any classical nonlinear optimization method. In this work, we seek the solution using the classical nonlinear least squares estimate method [28].

Moreover, c_U should be also satisfied with the constraint given by Eq. (8). Otherwise, this meta-class cannot be included in the credal classification results. In real applications, many unacceptable meta-classes will be eliminated through this step, and we just consider only the selected acceptable meta-classes as true focal elements of the bba. By doing so, we greatly reduce the computational complexity, which is very appealing for most engineering applications.

3.2. Construction of bba's

Let us consider a particular object $x_s \in X, s = 1, \dots, n$ to classify over the frame of discernment $\Omega = \{w_0, w_1, \dots, w_h\}$ using the framework of the belief functions. The mass of belief of the specific class (e.g. w_i) should depend on the Mahalanobis distance between the object and the corresponding center of class, and the bigger distance generally leads to the smaller mass of belief. If x_s is more close to a specific class center (e.g. c_i), it indicates that x_s belongs very likely to the class w_i as done in the classical way. So the initial mass of x_s of 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, which is denoted as

$$\tilde{m}(w_i) = f_1(d(x_s, c_i)), \quad \forall i = 1, \dots, h \quad (10)$$

The credal classification approach offers the possibility that the object belongs with different masses of belief to all the specific classes, and also to some meta-classes as well. The meta-classes are introduced as a means for modeling the imprecision of the class of the object. To reduce the computational burden, we have shown in the previous step devoted to the determination of meta-class center how some unacceptable meta-classes can be reasonably ignored. Moreover, we can reasonably assume that the object is close to the true class it belongs to in general. Consequently, the object should not very likely belong to the classes very far away of its true class. Based on this remark, we also consider (for the construction of the bba) the compatibility of the meta-classes according to the ascending order of the distances between the object and all the centers of specific classes.

The specific classes are listed in the ascending order of the distances of x_s to the centers as $(w_i, w_j, w_k, \dots, w_g)$. It means that x_s belongs most likely to w_i , then to w_j, w_k, \dots, w_g . Thus, we just need to consider only the nested meta-classes $w_i \cup w_j, w_i \cup w_j \cup w_k, \dots, w_i \cup w_j \cup w_k \dots \cup w_g$ because the object x_s will not very likely belong to the other meta-classes.

For example, let us consider the simple frame $\Omega = \{w_1, w_2, w_3\}$, and an object x_s which is the most close to w_1 , and then to w_2 , and then to w_3 . In that case, we select only the following nested meta-classes $w_1 \cup w_2$ and $w_1 \cup w_2 \cup w_3$ as potential focal elements. The meta-class $w_2 \cup w_3$ is not considered as compatible because the true class of x_s cannot be reasonably compatible with $w_2 \cup w_3$ only (because x_s is in fact the most close to w_1). Moreover, if some selected compatible nested classes appear finally unacceptable (according to the step of meta-class center determination described in Section 3.1), they will be ignored in the construction of the bba for the credal classification of the object.

Once all the acceptable meta-classes of the object x_s have been determined, we can proceed with the computation of the mass of these meta-classes (i.e. the focal elements of cardinalities greater than one). The principle of construction of the mass for a meta-class U is based on the following considerations:

- If an object is committed to a meta-class U , then of course it should be very close to the center c_U of this meta-class.
- the ratio $\gamma = d_{\max}/d_{\min}$ of the maximum distance d_{\max} of the object to the centers of the specific classes included in U over the minimum distance d_{\min} is used to measure the degree of distinguishability among the classes in U . The smaller ratio indicates a poor distinguishability degree among the classes in U from the object. The object committed to a meta-class must have a small ratio value (close to one) indicating that the involved specific classes are not very distinguishable for the object. So the value of the ratio γ will be used to put more or less mass of belief to the meta-class U .

⁷ Then U is a focal element of the bba.

Based on these considerations, the mass of belief of assignment of the object \mathbf{x}_s with the meta-class U is mathematically defined as

$$\tilde{m}(U) = f_2(d(\mathbf{x}_s, \mathbf{c}_U), \gamma_U) \quad (11)$$

where

$$d(\mathbf{x}_s, \mathbf{c}_U) = \frac{1}{|U|} \sum_{w_i \in U} \sqrt{\sum_{k=1}^N \frac{(\mathbf{x}_s(k) - \mathbf{c}_U(k))^2}{\delta_i(k)^2}} \quad (12)$$

$$\gamma_U = \frac{\max_{w_i \in U} d(\mathbf{x}_s, \mathbf{c}_i)}{\min_{w_i \in U} d(\mathbf{x}_s, \mathbf{c}_i)} \quad (13)$$

The smaller value of $d(\mathbf{x}_s, \mathbf{c}_U)$ and γ_U will yield bigger mass of belief $\tilde{m}(U)$, and vice versa. Hence, $f_2(\cdot)$ should be a monotone decreasing function with respect to $d(\mathbf{x}_s, \mathbf{c}_U)$ and γ_U .

To get good results, the functions $f_1(\cdot)$ and $f_2(\cdot)$ must be determined according to the application under concern. Unfortunately, we do not find yet general guidelines for the selection of these functions. This topic of research is actually under investigations. In this work, we have chosen the exponential decreasing function because it is commonly used in many engineering applications [13,18].

In summary, the (unnormalized) masses of belief for the specific classes and the acceptable meta-classes are finally given by

$$\tilde{m}(w_i) = e^{-d(\mathbf{x}_s, \mathbf{c}_i)} \quad (14)$$

$$\tilde{m}(U) = e^{-\lambda_U \gamma_U d(\mathbf{x}_s, \mathbf{c}_U)} \quad \text{for } |U| \geq 2 \quad (15)$$

with $\lambda_U = \eta |U|^\alpha$. The quantity $|U|^\alpha$ is a penalized parameter for the meta-classes having a big cardinality value. In most cases, the classification of the object is imprecise only among a small number of specific classes, and there are usually only few objects to assign with the meta-classes having big cardinalities. Thus, bigger cardinalities generate stronger penalization. η is a tuning parameter used to control the number of objects committed to the meta-classes. In practice, we always have to find a good compromise between the error rate and the imprecision rate. The guidelines for tuning the parameters are given at the end of this section.

The outlier class w_0 is also taken into account to deal with the case where the potential outliers (noise) can be involved. The object will be considered as outlier if it is far from all the other classes according to a given outlier threshold t . The mass of the object in the outlier class is defined by

$$\tilde{m}(w_0) = e^{-t} \quad (16)$$

$\forall A \subseteq \Omega$, the previous unnormalized mass of belief $\tilde{m}(\cdot)$ is normalized as follows:

$$m(A) = \frac{\tilde{m}(A)}{\sum_{B \subseteq \Omega} \tilde{m}(B)} \quad (17)$$

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

Guidelines for tuning the parameters in the CCR approach: The parameters α , η and t can be optimized using the training data with the cross-validation method (e.g. leave-one-out) before the application of CCR. The bigger penalized parameter α will lead to smaller number of the objects in the meta-class with big cardinality, and the suitable value can be found according to the classification results of the training data. Generally, one can take $\alpha \in [1, 3]$, e.g. 1 or 2. The parameter η is used to control the number of objects in the meta-classes. The bigger value of η will produce smaller number of objects committed to the meta-classes. It is recommended to take $\eta \in (0, 1)$, but the exact value of η can be tuned according to the imprecision degree (i.e. the rate of the objects in the meta-classes) of the classification results one can accept in the application under concern. The outlier threshold t

should be determined according to the outlier rate one expects in the classification. The bigger t will cause smaller number of outliers, and we generally recommend to take $t \in [2, 5]$.

4. Evaluation of CCR on artificial and real data sets

In this section we present four experiments to evaluate and compare the performances of CCR with respect to four classical methods: (1) the Classification And Regression Tree (CART) [29], (2) the Artificial Neural Networks (ANN) [30], (3) the EK-NN [13], and (4) the BK-NN [14]. The experiment #1, based on artificial data sets, is presented to show the difference between the credal classification and the classical methods. The experiment #2 allows to evaluate the performance of CCR with respect to the other methods based on a 4-class artificial data set. The experiment #3 is used to illustrate the efficiency of CCR for dealing with the large scale data set. The experiment #4 is based on four real-data sets from UCI [31]. It shows the advantage of CCR over the other methods. The different methods in the experiments have been programmed and tested with MatlabTM software.

In order to show the ability of CCR to deal with the meta-classes, the class of each object is decided according to the maximal mass of belief criterion. In CCR, the imprecise objects do not count really as misclassifications, but as imprecise classifications in fact. In our simulations, a misclassification is declared (counted) as soon as an object truly originated from w_i is classified into A with $w_i \cap A = \emptyset$. If $w_i \cap A \neq \emptyset$ and $A \neq w_i$ then it will be considered as an imprecise classification. In our experiments, the error rate and imprecision rate are introduced to evaluate the performance of CCR. The error rate R_e is calculated by $R_e = (N_e/T) \times 100\%$, where N_e is number of misclassification errors, and T is the number of objects under test. The imprecision rate R_{ij} is calculated by $R_{ij} = (N_{ij}/T) \times 100\%$, where N_{ij} is number of objects committed to the meta-classes with the cardinality value j .

4.1. Experiment #1 (with artificial data sets)

4.1.1. Test #1: A 2-class problem with artificial data

Two classes of artificial data set w_1 and w_2 are obtained from two uniform distributions as shown in Fig. 2(a). Each class has 200 training samples and 200 test samples. The uniform distributions of the samples of the two classes are characterized by the following bounds:

	x-label interval	y-label interval
w_1	(4, 6)	(0, 10)
w_2	(0, 10)	(4, 6)

Three classical classifiers (CART, ANN and EK-NN) are compared with the proposed CCR method. A particular value of $K=9$ is selected here for EK-NN, since it provides good results. The other parameters in EK-NN are optimized using the method introduced in [32]. In ANN, we used the feed-forward back propagation network with $epochs=500$ and $goal=0.001$ in all the experiments. In CCR, one has taken $\alpha=1$, $t=2$, and tested two different values of η to show its influence on the performances of CCR. The classification results of the objects with the different methods are given in Fig. 2(b)–(f). For notation conciseness, we have denoted $w^{te} \triangleq w^{test}$, $w^{tr} \triangleq w^{training}$ and $w_{i,\dots,k} \triangleq w_i \cup \dots \cup w_k$.

The misclassification rate obtained by the different methods is indicated in the title of each subfigure. The objects of classes w_1 and in w_2 are distributed over two overlapping areas following

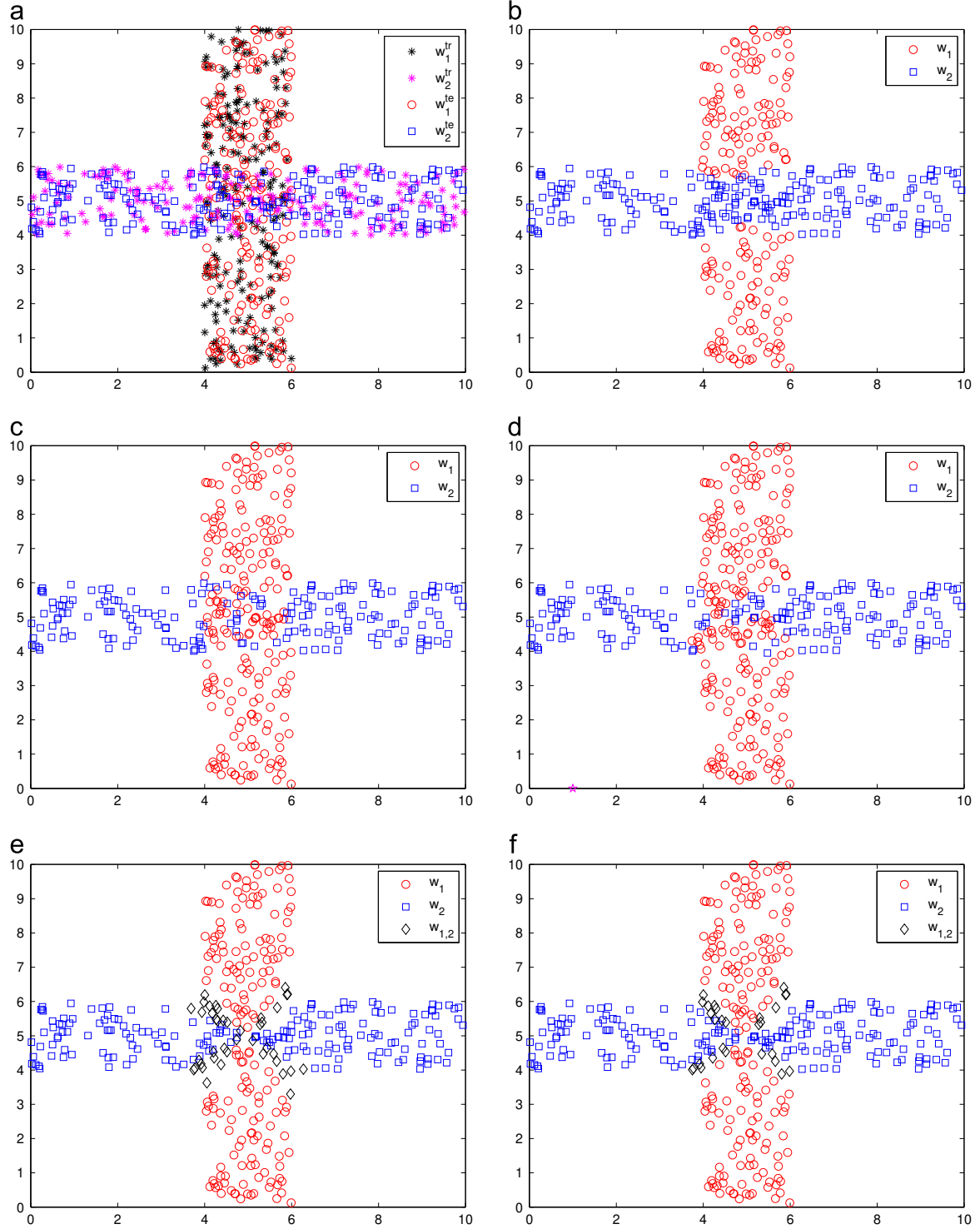


Fig. 2. Classification results obtained by different methods for a 2-class problem. (a) Original data, (b) classification result by ANN ($R_e = 9.00$), (c) classification result by CART ($R_e = 10.75$), (d) classification result by EK-NN ($R_e = 12.25$), (e) classification result by CCR with $\eta = 0.25$ ($R_e = 4.25, R_{I_2} = 10.25$) and (f) classification result by CCR with $\eta = 0.3$ ($R_e = 5.25, R_{I_2} = 7.00$).

a cross shape as shown in Fig. 2(a). Obviously, all objects belonging to the middle of the cross area are really difficult to associate with a particular class. However, EK-NN, CART and ANN just commit these objects into a specific class w_1 or w_2 as shown in Fig. 2(b)–(d). Such classification methods generate many misclassification errors (the error rate is about ten percent). CCR provides one more meta-class $w_1 \cup w_2$ as shown in Fig. 2(e) and (f). The classes w_1 and w_2 are undistinguishable for all the objects located in the intersecting

(overlapping) zone. Thus, it is more judicious and prudent to assign these objects to the meta-class $w_1 \cup w_2$. By doing this, one greatly reduces the number of misclassification, and also deeply reveals the imprecision degree of class of the objects. Once the tuning parameter $\eta = 0.25$ increases to $\eta = 0.3$, the imprecision rate will decrease but meanwhile the error rate will increase. So one should find a good compromise between the error rate and imprecision rate by tuning η in the training data space.

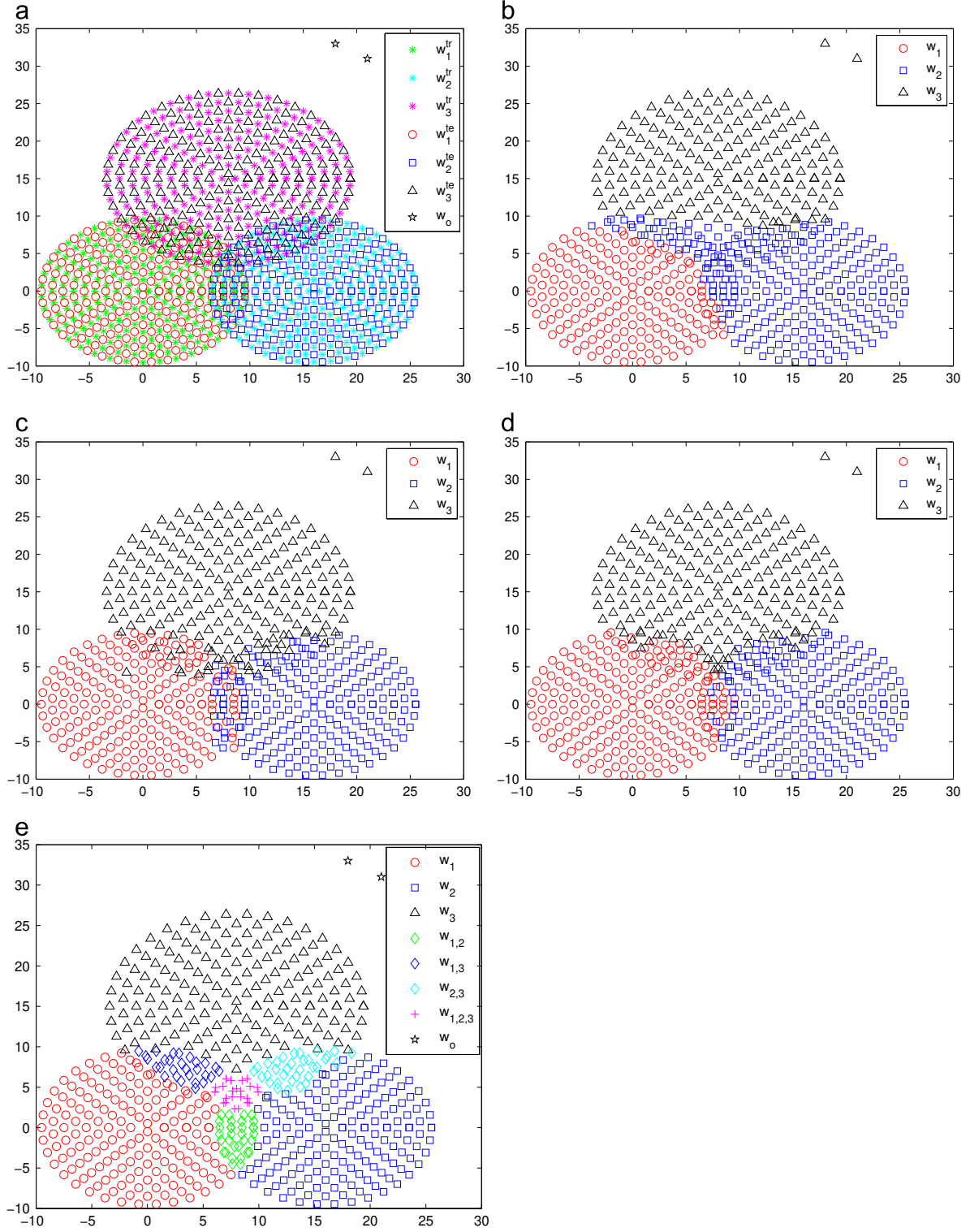


Fig. 3. Classification results by different methods for a 3-class problem. (a) Original data, (b) classification result by ANN ($R_e=14.42$), (c) classification result by CART ($R_e=10.86$), (d) classification result by EK-NN ($R_e=12.68$) and (e) classification result by CCR with $\eta=0.4$ ($R_e=1.13, R_{t_2}=15.07, R_{t_3}=3.73$).

4.1.2. Test #2: A 3-class problem with artificial data

In this second test, we consider a particular 3-classes data set in a round shape as shown in Fig. 3(a). This data set consists of 615 training data points and 617 test data points including two noisy data (outliers). The radii of the circles for w_1, w_2 and w_3 are $r_1=10, r_2=10, r_3=12$. The centers of three circles are located at $\mathbf{c}_1=(0,0)$, $\mathbf{c}_2=(16,0)$ and $\mathbf{c}_3=(8,15)$. CCR is applied for the classification of this particular data set and it is compared with

the CART, ANN and EK-NN classification methods. A particular value of $K=9$ is also selected in EK-NN. In CCR, we have chosen the tuning parameters $\alpha=2, t=2$ and $\eta=0.4$. The classification results obtained by the different methods are shown in Fig. 3(b)–(e).

The error rate and the imprecision rate of classification results obtained by the different methods are also given in the title of each subfigure. In Fig. 3(a), one sees that the three classes w_1, w_2 and w_3 partly overlap on their borders, and the points belonging to

Table 1

Classification results of different methods (in %).

Number of samples	ANN (R_e, t)	CART (R_e, t)	EK-NN (R_e, t)	BK-NN (R_e, R_{I_j}, t)	CCR (R_e, R_{I_j}, t)
$N=200$	(13.40, 10.8327)	(14.17, 0.0265)	(11.20, 0.5171)	(9.06, 5.39, 2.3816)	(8.45, 6.53, 0.0125)
$N=300$	(12.73, 11.8592)	(14.23, 0.0374)	(11.02, 0.6518)	(8.89, 3.17, 3.4892)	(8.06, 6.20, 0.0140)
$N=500$	(13.08, 13.0573)	(14.55, 0.0546)	(10.97, 1.1950)	(8.28, 4.06, 5.6056)	(7.50, 6.12, 0.0156)

Table 2

Classification results of different methods (in %).

Number of samples	ANN (R_e, t)	CART (R_e, t)	EK-NN (R_e, t)	CCR (R_e, R_{I_j}, t)
$N=8000$	(33.09, 15.6313)	(29.59, 1.2168)	(8.46, 47.5023)	(5.26, 5.84, 0.2340)
$N=40\ 000$	(35.04, 58.9684)	(26.66, 6.4428)	(8.25, 1669.1)	(5.15, 6.41, 1.1544)
$N=200\ 000$	(33.93, 241.7703)	(24.34, 35.1470)	NA	(5.11, 6.24, 5.8032)
$N=1\ 000\ 000$	NA	(22.25, 200.3053)	NA	(5.14, 6.16, 29.0162)

the overlapped zones are really difficult to classify correctly due to their ambiguity. Moreover, two noisy points far from the other data are included in the test data set. As shown in Fig. 3(b)–(d), ANN, CART and EK-NN produce only three singleton clusters w_1 , w_2 and w_3 respectively. Thus, most of the points in the overlapped zone are probably misclassified because of the inherent limitation of the framework adopted for these methods. These classifiers cannot detect the noisy data (outliers), and they all commit the noisy data into the class w_3 . 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 and w_1 and w_3 are respectively committed to $w_1 \cup w_2$, $w_2 \cup w_3$ and $w_1 \cup w_3$ as shown in Fig. 3(e) because these points are really difficult to classify correctly into a particular class. All of the three classes overlap in their middle, and the points in this zone are prudently committed to the meta-class $w_1 \cup w_2 \cup w_3$ because their classes are totally imprecise with respect to w_1 , w_2 and w_3 . CCR is also able to well detect the outliers. This example clearly shows the potential interest of the credal classification done by this new CCR approach.

4.2. Experiment #2 (with artificial data sets)

In this second experiment, the statistics of the performances of CCR are compared with CART, ANN, EK-NN and BK-NN on a 4-class artificial data set, which is generated from four 2D Gaussian distributions characterizing the classes w_1 , w_2 , w_3 and w_4 with the following means vectors

$$\begin{aligned}\mu_1 &= (0, 0), \Sigma_1 = 2 \cdot \mathbf{I} \\ \mu_2 &= (7, 0), \Sigma_2 = 2.5 \cdot \mathbf{I} \\ \mu_3 &= (15, 0), \Sigma_3 = 3 \cdot \mathbf{I} \\ \mu_4 &= (22, 0), \Sigma_4 = 2 \cdot \mathbf{I}\end{aligned}$$

There are 3×200 test objects, and the training sets contain $3 \times N$ samples (for $N=200, 300, 500$).

For EK-NN and BK-NN methods, the values of K ranging from 5 to 15 neighbors have been tested. The error rates R_e , the imprecision rates R_{I_j} , and the computation time t (in seconds) have been averaged over 10 Monte Carlo runs (i.e. 10 independent random generation of the data sets). The results obtained with the different classifiers are shown in Table 1. The BK-NN and CCR have been tuned to get a good compromise between the misclassification error and the imprecision of the results.

The meta-classes with cardinalities bigger than two are not considered in this application, that is why we did just mention R_{I_2} in Table 1. One sees in Table 1 that CCR and BK-NN produce smaller error rate than other methods. This is normal because the

objects that are difficult to classify correctly have been assigned to the associated meta-classes. In general, the classification results of BK-NN and CCR are similar. The error rate for CCR is a bit lower than for BK-NN, but in counterpart the imprecision rate for BK-NN is lower than for CCR. However, BK-NN requires much more computational time than CCR, which shows that the computational burden of BK-NN is much bigger than CCR. CCR consumes much less time than any other tested methods which indicates that CCR has the least computational complexity which offers a strong advantage for some engineering applications with respect to other methods.

4.3. Experiment #3 (with large scale artificial data sets)

The performance of CCR for dealing with large scale data sets (i.e. big number of samples with high-dimensional features) is evaluated in this experiment by comparing CCR with several other classical methods⁸ (ANN, CART and EK-NN).

In this experiment, an artificial data set with four class w_1 , w_2 , w_3 and w_4 is generated from four 30D Gaussian distributions with the means vectors and covariance matrices as follows:

$$\begin{aligned}\mu_1 &= \mathbf{zeros}(1, 30), \Sigma_1 = 10 \cdot \mathbf{I} \\ \mu_2 &= 5 \cdot \mathbf{ones}(1, 30), \Sigma_2 = 10 \cdot \mathbf{I} \\ \mu_3 &= 20 \cdot \mathbf{ones}(1, 30), \Sigma_3 = 15 \cdot \mathbf{I} \\ \mu_4 &= 30 \cdot \mathbf{ones}(1, 30), \Sigma_4 = 15 \cdot \mathbf{I}\end{aligned}$$

Here $\mathbf{zeros}(1, 30)$ represents the 30-dimensional vector with value of zero in each dimension, and $\mathbf{ones}(1, 30)$ is the 30-dimensional vector with value of one in each dimension, and \mathbf{I} denotes the 30×30 identity matrix.

In each class, we use the same number (i.e. n) of training samples and test samples. So there are totally $N = 4 \times n$ training samples and $N = 4 \times n$ test samples, and we take $N=8000, 40\ 000, 200\ 000, 1\ 000\ 000$. In EK-NN, the values of K ranging from 5 to 15 neighbors are tested. It has have been tuned to get a good compromise between the misclassification error and the imprecision of the results by CCR. The error rates R_e , the imprecision rates R_{I_j} , and the computation time t (in seconds) are the average value over 10 Monte Carlo runs. The results produced by the different classifiers are illustrated in Table 2. In Table 2, 'NA' means 'Not Applicable'.

⁸ It is well known that the K-NN based methods (e.g. EK-NN, BK-NN, etc) are usually not very effective for dealing with the big data set due to the large computation burden. We have shown that BK-NN can produce results similar to CCR, but it requires more computational time than CCR, EK-NN and CART. So we just use the EK-NN method here to compare its performance with CCR.

Table 3
Main information of the real data sets used in our tests.

Name	Classes	Attributes	Instances
Iris	3	4	150
Seeds	3	7	210
Wine	3	7	255
Yeast	3	9	683

Table 4
Classification results of different methods (in%).

Data sets	ANN (R_e, t)	CART (R_e, t)	EK-NN (R_e, t)	BK-NN (R_e, R_{I_2}, t)	CCR (R_e, R_{I_2}, t)
Iris	(4.00, 4.5708)	(5.33, 0.0156)	(3.98, 0.0094)	(2.12, 6.67, 0.0581)	(2.00, 6.67, 0.0000)
Seeds	(16.665, 9.3289)	(11.90, 0.0234)	(10.57, 0.0296)	(7.71, 4.16, 0.0971)	(7.14, 6.19, 0.0000)
Wine	(33.71, 8.4553)	(8.89, 0.0312)	(28.94, 0.0135)	(22.11, 11.80, 0.0737)	(3.37, 0, 0.0078)
Yeast	(51.42, 9.9841)	(37.73, 0.0936)	(36.51, 0.2366)	(27.76, 14.42, 2.4350)	(25.31, 19.43, 0.0156)

We can see that CCR produces the lowest error rate with some partial imprecision results, since it assigns some objects that are hard to be correctly classified into the proper meta-classes. Meanwhile, CCR consumes the shortest operation time. EK-NN can obtain the reasonable classification results, but it requires the longest running time, which is the main drawback of the K-NN based methods. EK-NN is even not applicable when the number of samples is big (i.e. $N=200\ 000$ and $N=1\ 000\ 000$), since it takes too long time, which is not convenient in many cases where the high speed of execution is necessary. ANN and CART cause much higher error rate than CCR and EK-NN, and they are also much more time-consuming than CCR. ANN is not applicable for the big data set (i.e. $N=1\ 000\ 000$) because of its high computational burden. So it indicates that CCR is effective for dealing with the large scale data set thanks to its low computational and complexity burden.

4.4. Experiment #4 (with real data sets)

Four well-known real data sets obtained from UCI Machine Learning Repository [31] (the Iris, Seeds, Wine and Yeast data sets) have been tested in this experiment to evaluate the performances of CCR compared with CART, ANN, EK-NN and BK-NN. For the Yeast data set, three classes named as CYT, NUC and ME3 are selected here, since these three classes are close and difficult to discriminate. The main characteristics of the four data sets are summarized in Table 3 below. All the detailed information can be found on UCI repository archive at <http://archive.ics.uci.edu/ml/>.

The k -fold cross validation is performed on the four data sets by different classification methods, and k generally remains a free parameter [33]. We use the simplest 2-fold cross validation⁹ here, since it has the advantage that the training and test sets are both large, and each sample is used for both training and testing on each fold. The tuning parameter of CCR and BK-NN were optimized using the training samples. The classification results including R_e and R_{I_j} of BK-NN and EK-NN are calculated with values of K ranging from 5 to 15. The reported error rates R_e , the imprecision rates R_{I_j} , and the computation time t (in seconds) for the different methods are given in Table 4.

In these tests, no object is committed to the meta-class with cardinality value of three, and that is why we have just given R_{I_2} in Table 4. From Table 4, one sees that CCR and BK-NN produce the

smaller error rate than other classical methods. It is normal because the objects difficult to classify correctly have been reasonably and automatically committed to the associated meta-classes. It shows that the credal classification can effectively reduce error occurrences, and the meta-classes indicate that the attributes information is not good enough to obtain the correct specific class of some objects. In that case, some other complementary sources of information, or techniques, will be necessary if one wants to precisely discriminate the objects committed to the meta-classes with high belief mass value (if a precise classification is absolutely required). The CCR and BK-NN methods provide similar performances for the Iris, Seeds and Yeast data sets according to the compromise between error rate and imprecision rate. For the Wine data set, CCR yields the lowest error rate due to its inherent working principle which is very different of the other classifiers. It is worth noting that BK-NN requires a very long running time due to the heavy computational load. The proposed CCR method requires less computational time than the other methods. This shows again that CCR working with credal classification can deal efficiently with uncertain data using belief functions with a serious computational complexity advantage over other methods.

5. Conclusions

A new simple and effective credal classification rule (CCR) based on the belief functions has been presented in this work to deal with the classification of uncertain data. CCR strengthens the robustness of results by reducing the misclassification errors thanks to the introduction of meta-classes. The CCR approach is also able to detect the outliers in the data sets. In CCR, each specific class corresponds to a center (i.e. prototype) obtained using the training data, and the center of meta-class is located at the equal Mahalanobis distances to all the centers of the involved specific classes. Mahalanobis distance is used here to deal with the anisotropic data sets. The acceptable meta-classes are selected according to the current context and distance ratios, and all the unacceptable meta-classes are automatically rejected to reduce the number of focal elements and the computational complexity. A tuning parameter has been introduced in CCR to control the number of objects in the meta-classes. The output of CCR can be used efficiently to alert the classification system designer that other complementary information sources are necessary to remove (or reduce) the ambiguity of the classification of some particular data points. Several experiments using both the artificial and real data sets have been presented to evaluate the performance

⁹ More precisely, the samples in each classes are randomly assigned to two sets S_1 and S_2 having equal size. Then we train on S_1 and test on S_2 , and reciprocally.

of CCR with respect to other methods. Our results show that CCR is able to provide good credal classification results with a relatively low computational complexity with respect to other methods. Belief functions provide an effective tool to model the uncertainty and imprecision of data. It is possible to incorporate belief functions into other methods (e.g. CART, ANN, etc) to improve the classification performance, and we will make more investigations on this topic in our future research work.

Conflict of interest statement

None declared.

Acknowledgements

This work has been partially supported by National Natural Science Foundation of China (Nos. 61135001, 61374159).

References

- [1] A.-L. Jousselme, P. Maupin, E. Bossé, Uncertainty in a situation analysis perspective, in: *Proceedings of the 6th International Conference on Information Fusion*, Queensland, Australia, July 2003.
- [2] A.-L. Jousselme, C. Liu, D. Grenier, E. Bossé, Measuring ambiguity in the evidence theory, *IEEE Trans. SMC: Part A* 36 (September (5)) (2006) 890–903 (see also comments in [3]).
- [3] G. Klir, H. Lewis, Remarks on “Measuring ambiguity in the evidence theory”, *IEEE Trans. SMC: Part A* 38 (July (4)) (2008) 895–999.
- [4] G. Shafer, *A Mathematical Theory of Evidence*, Princeton University Press, Princeton, 1976.
- [5] F. Smarandache, J. Dezert (Eds.), *Advances and Applications of DSMT for Information Fusion*, American Research Press, Rehoboth, vol. 1–3, 2004–2009, see (<http://fs.gallup.unm.edu/DSMT.htm>) or (<http://www.onera.fr/staff/jean-dezert?page=2>).
- [6] P. Smets, The combination of evidence in the transferable belief model, *IEEE Trans. Pattern Anal. Mach. Intell.* 12 (5) (1990) 447–458.
- [7] P. Smets, R. Kennes, The transferable belief model, *Artif. Intell.* 66 (2) (1994) 191–243.
- [8] P. Smets, Analyzing the combination of conflicting belief functions, *Inf. Fus.* 8 (4) (2007) 387–412.
- [9] M.-H. Masson, T. Denœux, ECM: an evidential version of the fuzzy c-means algorithm, *Pattern Recognit.* 41 (4) (2008) 1384–1397.
- [10] T. Denœux, M.-H. Masson, EVCLUS: Evidential CLUSTERing of proximity data, *IEEE Trans. Syst., Man Cybern. Part B* 34 (1) (2004) 95–109.
- [11] Z.-g. Liu, J. Dezert, G. Mercier, Q. Pan, Belief C-Means: an extension of fuzzy c-means algorithm in belief functions framework, *Pattern Recognit. Lett.* 33 (3) (2012) 291–300.
- [12] H. Laanaya, A. Martin, D. Aboutajdine, A. Khenchaf, Support vector regression of membership functions and belief functions – application for pattern recognition, *Inf. Fus.* 11 (4) (2010) 338–350.
- [13] T. Denœux, A k-nearest neighbor classification rule based on Dempster–Shafer Theory, *IEEE Trans. Syst., Man Cybern.* 25 (05) (1995) 804–813.
- [14] Z.-g. Liu, Q. Pan, J. Dezert, A new belief-based K-nearest neighbor classification method, *Pattern Recognit.* 46 (3) (2013) 834–844.
- [15] T. Denœux, P. Smets, Classification using belief functions: relationship between case-based and model-based approaches, *IEEE Trans. Syst., Man Cybern., Part B* 36 (6) (2006) 1395–1406.
- [16] F. Karem, M. Dhibi, A. Martin, Combination of supervised and unsupervised classification using the theory of belief functions, in: T. Denœux, M.-H. Masson (Eds.), *Belief Functions: Theory and Applications*, vol. 164, AISC, Springer Heidelberg New York Dordrecht London, 2012, pp. 85–92.
- [17] T. Denœux, Z. Youne, F. Abdallah, Representing uncertainty on set-valued variables using belief functions, *Artif. Intell.* 174 (7–8) (2010) 479–499.
- [18] J. Dezert, Z.-g. Liu, G. Mercier, Edge detection in color images based on DSMT, in: *Proceedings of the 14th International Conference on Information Fusion*, Chicago, USA, July 2011.
- [19] Z.-g. Liu, J. Dezert, G. Mercier, Q. Pan, Dynamic evidential reasoning for change detection in remote sensing images, *IEEE Trans. Geosci. Remote Sens.* 50 (5) (2012) 1955–1967.
- [20] Z.-g. Liu, J. Dezert, Q. Pan, G. Mercier, Combination of sources of evidence with different discounting factors based on a new dissimilarity measure, *Decis. Support Syst.* 52 (2011) 133–141.
- [21] A. Martin, C. Osswald, J. Dezert, F. Smarandache, General combination rules for qualitative and quantitative beliefs, *J. Adv. Inf. Fus.* 3 (2) (2008) 67–89.
- [22] T. Denœux, Conjunctive and disjunctive combination of belief functions induced by nondistinct bodies of evidence, *Artif. Intell.* 172 (2–3) (2008) 234–264.
- [23] J.C. Bezdek, *Pattern Recognition with Fuzzy Objective Function Algorithms*, Plenum Press, New York, 1981.
- [24] R.N. Dave, Clustering relational data containing noise and outliers, *Pattern Recognit. Lett.* 12 (1991) 657–664.
- [25] T. Denœux, Maximum likelihood estimation from uncertain data in the belief function framework, *IEEE Trans. Knowl. Data Eng.* 25 (1) (2013) 119–130.
- [26] T. Denœux, A neural network classifier based on Dempster–Shafer theory, *IEEE Trans. Syst., Man Cybern. A* 30 (2) (2000) 131–150.
- [27] B.W. Silvean, M.C. Jones, E. Fix, J.L. Hodges, 1951. An important contribution to nonparametric discriminant analysis and density estimation, *Int. Stat. Rev.* 57 (December (3)) (1989) 227–233.
- [28] T.F. Coleman, *Large Sparse Numerical Optimization*, Springer-Verlag, New York, 1984.
- [29] L. Breiman, J. Friedman, C.J. Stone, R.A. Olshen, *Classification and Regression Trees*, Chapman and Hall (Wadsworth, Inc.), New York, 1984.
- [30] C.M. Bishop, *Neural Networks for Pattern Recognition*, Oxford University Press, Oxford, 1995.
- [31] A. Frank, A. Asuncion, *UCI Machine Learning Repository*, University of California, School of Information and Computer Science, Irvine, CA, USA, 2010 (<http://archive.ics.uci.edu/ml>).
- [32] L.M. Zouhal, T. Denœux, An evidence-theoretic k-NN rule with parameter optimization, *IEEE Trans. Syst., Man Cybern., Part C* 28 (2) (1998) 263–271.
- [33] S. Geisser, *Predictive Inference: An Introduction*, Chapman and Hall, New York, NY, 1993.