Pattern Recognition 47 (2014) 2532–2541

Contents lists available at ScienceDirect

Pattern Recognition

journal homepage: www.elsevier.com/locate/pr

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

ARTICLE INFO

Article history:
Received 24 May 2013
Received in revised form
23 September 2013
Accepted 23 January 2014
Available online 3 February 2014

Keywords:
Credal classification
Data classification
Belief functions
Evidence theory
Uncertain data

ABSTRACT

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-classes[1] (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.

0031-3203/$ - see front matter
http://dx.doi.org/10.1016/j.patcog.2014.01.011

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 $\Theta=\{\theta_1,\theta_2,\ldots,\theta_h\}$ of $h>1$ mutually exclusive and exhaustive hypotheses θ_i, $i=1,2,\ldots,h$. 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

$$2^\Theta=\{A|A\subseteq\Theta\} \tag{1}$$

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

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 elements[2] 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

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

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 $M=(\mathbf{m_1},\ldots,\mathbf{m_n})$, 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.

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

$$\begin{cases} Bel(X)=\sum\limits_{Y\in 2^\Theta|Y\subseteq X} m(Y)\\ Pl(X)=\sum\limits_{Y\in 2^\Theta|X\cap Y\neq\varnothing} m(Y)\end{cases} \tag{3}$$

$Bel(X)$ 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 $\sum_{\theta_i\in\Theta}Bel(\theta_i)\le 1$. $Pl(X)$ 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 $\forall A\in 2^\Theta\backslash\{\varnothing\}$ by

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

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, $\theta_1\cup\theta_2\cup\ldots\cup\theta_h\equiv\Theta$.