Neurocomputing 133 (2014) 459–470

Contents lists available at ScienceDirect

Neurocomputing

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

Classification of uncertain and imprecise data based on evidence theory

Zhun-ga Liu [a,*], Quan Pan [a], Jean Dezert [b]

[a] School of Automation, Northwestern Polytechnical University, Xi'an, China
[b] ONERA – The French Aerospace Lab, F-91761 Palaiseau, France

ARTICLE INFO

Article history:
Received 1 March 2013
Received in revised form
3 October 2013
Accepted 6 December 2013
Communicated by N.T. Nguyen
Available online 11 January 2014

Keywords:
Evidence theory
Credal classification
Belief functions
K-NN
Data classification

ABSTRACT

In this paper, we present a new belief $c \times K$ neighbor (BCKN) classifier based on evidence theory for data classification when the available attribute information appears insufficient to correctly classify objects in specific classes. In BCKN, the query object is classified according to its K nearest neighbors in each class, and $c \times K$ neighbors are involved in the BCKN approach (c being the number of classes). BCKN works with the credal classification introduced in the belief function framework. It allows to commit, with different masses of belief, an object not only to a specific class, but also to a set of classes (called meta-class), or eventually to the ignorant class characterizing the outlier. The objects that lie in the overlapping zone of different classes cannot be reasonably committed to a particular class, and that is why such objects will be assigned to the associated meta-class defined by the union of these different classes. Such an approach allows to reduce the misclassification errors at the price of the detriment of the overall classification precision, which is usually preferable in some applications. The objects too far from the others will be naturally considered as outliers. The credal classification is interesting to explore the imprecision of class, and it can also provide a deeper insight into the data structure. The results of several experiments are given and analyzed to illustrate the potential of this new BCKN approach.

1. Introduction

In classification problems, the case-based classifier can be a good solution to classify the new input sample (the query object under test) using the collection of labeled (training) samples when the complete statistical knowledge regarding the conditional density functions is not available. The well known case-based classification methods, like K-nearest neighbor (K-NN) [1,2], decision trees [3,4], support vector machine (SVM) [5–7], artificial neural networks (ANNs) [8], have been developed essentially based on probability measure, or fuzzy number for dealing with the uncertain data. The samples are allowed to belong to different specific classes with different memberships, and the class with the biggest membership is usually chosen as final assignment of the object to a class (i.e. the decision-making).

In the classification of uncertain and imprecise data, the given attribute information can be insufficient for making a correct specific classification of the objects. For example, the attribute data from different classes can be partly overlapped sometimes. Such objects lying in the overlapping zone are in fact very difficult to classify correctly in a specific class, since the (partly) overlapped classes become undistinguishable. Moreover, some outliers (noisy data) can also be present in some applications. The probabilistic framework cannot well model and manage the imprecision of data. In fact, the probabilistic framework captures only the randomness aspect of the data, but not the fuzziness, nor imprecision which is another inherent aspect of information content [9,10].

The belief functions [11,13,12,14], introduced originally in the mathematical theory of evidence theory by Shafer in 1976, also known as Dempster-Shafer Theory (DST), offer a rigorous mathematical formalism to model uncertain and imprecise information produced by a source of evidence. This formalism has been already applied in many fields, including classification [15–20], clustering [21–23] and information fusion [24,25]. A recent concept called credal partition [22] has been introduced by Denœux and Masson based on the belief function for data clustering (unsupervised classification). The credal partition is an extension of the probabilistic partition based on a frame of discernment $\Omega = \{w_1, \ldots, w_c\}$ that allows the samples to belong not only to the specific classes (e.g. w_i) but also to a set of classes called a meta-class (e.g. $w_i \cup w_j$) with different masses of belief. The credal partition provides a deeper insight into the data structure as already reported in [21]. An evidential version of fuzzy c-means (FCM) clustering method [21] inspired by FCM [26] and Dave's Noise-Clustering algorithm [27] has been developed using credal partition to deal with imprecise data and outliers.

* Corresponding author.
E-mail addresses: liuzhunga@gmail.com (Z.-g. Liu), jean.dezert@onera.fr (J. Dezert).

0925-2312/$ - see front matter © 2014 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.neucom.2013.12.009

Some data classifiers have already been developed based on belief functions in the past. For instance, Smets [28] and Appriou [29] have proposed the model-based classifier based on the Generalized Bayes Theorem (GBT) [17]. GBT is an extension of Bayes theorem in Smets transferable belief model (TBM) [12,13]. Some case-based classifiers have also been proposed by Denœux [15,16]. Particularly, the evidential version of K-nearest neighbors (EK-NNs) method has been proposed in [15] based on DST, for working only with the specific classes and the extra ignorant class defined by the union of all the specific classes. An ensemble technique for the combination of evidential K-NN classifier based on DST has been proposed in [30] to improve the accuracy. A neural network classifier has also been developed in [16] under the belief functions framework that allows one extra ignorant class as possible output of this classifier.

The meta-class defined by the union of several specific classes (say $w_i \cup w_j$, $w_i \cup w_j \cup w_k$, etc.) is very important and useful to explore the partial imprecision inherent of the data set. However, it has not been considered completely in the existing evidential classifiers developed so far. In this work, we propose a new case-based belief classifier working with credal classification corresponding to the credal partition in data clustering, where both the meta-classes and the outlier class are taken into account to fully characterize the uncertainty and imprecision inherent in the data set. This is the innovation of this paper.

In this new method, the sample (the object to assign) is classified using its neighborhood of the training data space, and the K nearest neighbors in each class are used. A total of $c \times K$ (c being the number of classes) neighbors is used to classify the object. This new method is called a *belief* $c \times K$ *neighbors* (BCKNs) classifier. In BCKN, $c \times K$ basic belief assignments (bba's) will be constructed according to the distance between the object and its selected neighbors. A global fusion of these bba's is done to decide the class, or the meta-class to assign for the object. The credal classification of BCKN can produce specific class, meta-class and outlier class.

An object that is very close to a particular class of data will be committed to this specific class. An object too far from all the training samples will be naturally considered as an outlier (noise), which is helpful for the outlier detection in some applications. If the object is close to several specific classes (e.g. when lying in the overlapping zone of several different classes), then this object will be committed to the meta-class defined by the union of these specific classes. The meta-class reveals the imprecision in the classification of this object, and can also reduce the misclassifications. Of course the commitments are done in a soft manner thanks to the computation of proper basic belief masses as it will be explained in detail in Section 3. Such a credal classification (a classification based on soft assignments represented by belief functions) is very interesting in many applications, especially those related to defense and security (like in target classification and tracking) because it is generally preferable to get a more robust (and eventually partially imprecise) classification result that could be precisiated later with additional techniques or resources, than to obtain directly with high risk a wrong precise classification from which an erroneous fatal decision would be drawn. This is the main reason why we develop such a type of classifiers.

If some samples are committed to the meta-classes, it implies that the used attributes information for classification is insufficient to get the specific classification for these samples. Thus, the output of BCKN can be considered as an interesting source of information to be fused with some other available complementary information sources (when available) for getting more precise classification results in the multi-source information fusion systems. Of course, other sophisticated and generally more costly techniques, like those applied in the military applications, could also be used to classify more precisely the objects in the meta-classes. The use of such additional sophisticated techniques highly depends on the importance of the consequences of the decision to take. The objects in a meta-class are usually a small subset of the total data set. So the price for the specific classification of these objects invoking costly sophisticated techniques can be acceptable for only a limited number of objects, but not for the whole data set at the very beginning of the classification process. Thus, the BCKN method provides a way to select the objects (in meta-class) that need a particular attention which should be treated cautiously, as far as important decisions to take are under concern (like in a military targeting process by example).

This paper is organized as follows. The background on the belief functions is briefly introduced in the next section. The details of BCKN are presented in Section 3. Several experiments are given in Section 4 to show how BCKN performs with respect to other classical methods. Concluding remarks are given in the last section of this paper.

2. Background on belief functions

The belief functions have been introduced in 1976 by Shafer in his mathematical theory of evidence, known also as Dempster–Shafer theory (DST) [11,13,12,14] because Shafer uses Dempster's fusion rule for combining belief basic assignments. We consider a finite discrete set $\Omega = \{w_1, w_2, \ldots, w_c\}$. Ω of $c > 1$ mutually exclusive and exhaustive hypotheses, which is called the *frame of discernment* (FoD) of the problem under consideration. The power-set of Ω denoted by 2^Ω contains all the subsets of Ω. For example, if $\Omega = \{w_1, w_2, w_3\}$, then $2^\Omega = \{\emptyset, w_1, w_2, w_3, w_1 \cup w_2, w_1 \cup w_3, w_2 \cup w_3, \Omega\}$. The union $\theta_i \cup \theta_j = \{\theta_i, \theta_j\}$ 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". So that Ω represents the full ignorance (uncertainty).

Shafer [11] considers the subsets as propositions in the case we are concerned with the true value of some quantity w taking its possible values in Ω. Then the propositions $\mathcal{P}_w(A)$ of interest are those of the form[1]: $\mathcal{P}_w(A) \triangleq$ the true value of w is in a subset A of Ω. Any proposition $\mathcal{P}_w(A)$ is thus in one-to-one correspondence with the subset A of Ω. Such a correspondence is very useful since it translates the logical notions of conjunction $\wedge$, disjunction $\vee$, implication $\Rightarrow$ and negation $\neg$ into the set-theoretic notions of intersection $\cap$, union $\cup$, inclusion $\subset$ and complementation $c(\cdot)$. Indeed, if $\mathcal{P}_w(A)$ and $\mathcal{P}_w(B)$ are two propositions corresponding to subsets A and B of Ω, then the conjunction $\mathcal{P}_w(A) \wedge \mathcal{P}_w(B)$ corresponds to the intersection $A \cap B$ and the disjunction $\mathcal{P}_w(A) \vee \mathcal{P}_w(B)$ corresponds to the union $A \cup B$. A is a subset of B if and only if $\mathcal{P}_w(A) \Rightarrow \mathcal{P}_w(B)$ and A is the set-theoretic complement of B with respect to Ω (written $A = c_\Omega(B)$) if and only if $\mathcal{P}_w(A) = \neg\mathcal{P}_w(B)$. In other words, the following equivalences are then used between the operations on the subsets and on the propositions: (intersection $\equiv$ conjunction), (union $\equiv$ disjunction), (inclusion $\equiv$ implication) and (complementation $\equiv$ negation).

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

$$\begin{cases} \sum_{A \in 2^\Omega} m(A) = 1 \\ m(\emptyset) = 0 \end{cases} \tag{1}$$

The subsets A of Ω such that $m(A) > 0$ are called the *focal elements* of $m(\cdot)$, and the set of all its focal elements is called the *core* of $m(\cdot)$. If A is a singleton element corresponding to specific class, the quantity $m(A)$ can be interpreted as the exact belief committed to the class A. $m(A \cup B)$ reflects the imprecision (non-specificity or

[1] We use the symbol $\triangleq$ to mean *equals by definition*; the right-hand side of the equation is the definition of the left-hand side.