Pattern Recognition 52 (2016) 85–95

Contents lists available at ScienceDirect

Pattern Recognition

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

Adaptive imputation of missing values for incomplete pattern classification

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

[a] School of Automation, Northwestern Polytechnical University, Xi'an, China
[b] ONERA - The French Aerospace Lab, F-91761 Palaiseau, France
[c] IRISA, University of Rennes 1, Rue E. Branly, 22300 Lannion, France

ARTICLE INFO

Article history:
Received 1 June 2015
Received in revised form
29 September 2015
Accepted 1 October 2015
Available online 20 October 2015

Keywords:
Belief function
Classification
Missing values
SOM
K-NN

ABSTRACT

In classification of incomplete pattern, the missing values can either play a crucial role in the class determination, or have only little influence (or eventually none) on the classification results according to the context. We propose a credal classification method for incomplete pattern with adaptive imputation of missing values based on belief function theory. At first, we try to classify the object (incomplete pattern) based only on the available attribute values. As underlying principle, we assume that the missing information is not crucial for the classification if a specific class for the object can be found using only the available information. In this case, the object is committed to this particular class. However, if the object cannot be classified without ambiguity, it means that the missing values play a main role for achieving an accurate classification. In this case, the missing values will be imputed based on the *K*-nearest neighbor (K-NN) and Self-Organizing Map (SOM) techniques, and the edited pattern with the imputation is then classified. The (original or edited) pattern is classified according to each training class, and the classification results represented by basic belief assignments are fused with proper combination rules for making the credal classification. The object is allowed to belong with different masses of belief to the specific classes and meta-classes (which are particular disjunctions of several single classes). The credal classification captures well the uncertainty and imprecision of classification, and reduces effectively the rate of misclassifications thanks to the introduction of meta-classes. The effectiveness of the proposed method with respect to other classical methods is demonstrated based on several experiments using artificial and real data sets.

1. Introduction

In many practical classification problems, the available information for making object classification is partial (incomplete) because some attribute values can be missing due to various reasons (e.g. the failure or dysfunctioning of the sensors providing information, or partial observation of object of interest because of some occultation phenomenon). So it is crucial to develop efficient techniques to classify as best as possible the objects with missing attribute values (incomplete pattern), and the search for a solution of this problem remains an important research topic in the pattern classification field [1,2]. Some more details about pattern classification can be found in [3,4].

There have been many approaches developed for classifying the incomplete patterns [1], and they can be broadly grouped into four different types. The first (simplest) one is to remove directly the patterns with missing values, and the classifier is designed only for the complete patterns. This method is acceptable when the incomplete data set is only a very small subset (e.g. less than 5%) of the whole data set, but it cannot effectively classify the pattern with missing values. The second type is the model-based techniques [5]. The probability density function (PDF) of the input data (complete and incomplete cases) is estimated at first by means of some procedures, and then the object is classified using Bayesian reasoning. For instance, the expectation-maximization (EM) algorithm have been applied to many problems involving missing data for training Gaussian mixture models [5]. In the model-based methods, it must make assumptions about the joint distribution of all the variables in the model, but the suitable distributions sometimes are hard to obtain. The third type classifiers are designed to directly handle incomplete pattern without imputing the missing values, such as neural network ensemble methods [6], decision trees [7], fuzzy approaches [8] and support vector machine classifier [9]. The last type is the often used imputation (estimation) method. The missing values are filled with proper

* Corresponding author.
E-mail addresses: liuzhunga@nwpu.edu.cn (Z.-g. Liu), jean.dezert@onera.fr (J. Dezert), Arnaud.Martin@univ-rennes1.fr (A. Martin).

http://dx.doi.org/10.1016/j.patcog.2015.10.001
0031-3203/

estimations [10] at first, and then the edited patterns are classified using the normal classifier (for the complete pattern). The missing values and pattern classification are treated separately in these methods. Many works have been devoted to the imputation of missing data, and the imputation can be done either by the statistical methods, e.g. mean imputation [11] and regress imputation [2], or by machine learning methods, e.g. *K*-nearest neighbors imputation (KNNI) [12], Fuzzy *c*-means (FCM) imputation (FCMI) [13,14], and Self-Organizing Map imputation (SOMI) [15]. In KNNI, the missing values are estimated using *K*-nearest neighbors of object in training data space. In FCMI, the missing values are imputed according to the clustering centers of FCM and taking into account the distances of the object to these centers [13,14]. In SOMI [15], the best match node (unit) of incomplete pattern can be found ignoring the missing values, and the imputation of the missing values is computed based on the weights of the activation group of nodes including the best match node and its close neighbors. These existing methods usually attempt to classify the object into a particular class with maximal probability or likelihood measure. However, the estimation of missing values is in general quite uncertain, and the different imputations of missing values can yield very different classification results, which prevent us to correctly commit the object into a particular class.

Belief function theory (BFT), also called Dempster–Shafer theory (DST) [16] and its extension [18,17] offer a mathematical framework for modeling uncertainty and imprecise information [19]. BFT has already been applied successfully for object classification [20–28], clustering [29–33], multi-source information fusion [34–37], etc. Some classifiers for the complete pattern based on DST have been developed by Denœux and his collaborators to come up with the evidential *K*-nearest neighbors (EK-NN) [21], evidential neural network (ENN) [27], etc. The extra ignorance element represented by the disjunction of all the elements in the whole frame of discernment is introduced in these classifiers to capture the totally ignorant information. However, the partial imprecision, which is very important in the classification, is not well characterized. We have proposed credal classifiers [23,24] for complete pattern considering all the possible meta-classes (i.e. the particular disjunctions of several singleton classes) to model the partial imprecise information. The credal classification allows the objects to belong (with different masses of belief) not only to the singleton classes, but also to any set of classes corresponding to the meta-classes. In [23], a belief-based *K*-nearest neighbor classifier (BK-NN) has been presented, and the credal classification of object is done according to the distances between the object and its *K* nearest neighbors as well as two given (acceptance and rejection) distance thresholds. The K-NN classifier generally takes big computation burden, and this is not convenient for real application. Thus, a simple credal classification rule (CCR) [24] has been further developed, and the belief value of object associated with different classes (i.e. singleton classes and selected meta-classes) is directly calculated by the distance to the center of corresponding class and the distinguishability degree (w.r.t. object) of the singleton classes involved in the meta-class. The location of center of meta-class in CCR is considered with the same (similar) distance to all the involved singleton classes' centers. Moreover, when the training data is not available, we have also proposed several credal clustering methods [30–32] in different cases. Nevertheless, these previous credal classification methods are mainly for dealing with complete pattern without taking into account the missing values.

In our recent work, a prototype-based credal classification (PCC) [25] method for the incomplete patterns has been introduced to capture the imprecise information caused by the missing values. The object hard to correctly classify is committed to a suitable meta-class by PCC, which well characterizes the imprecision of classification due to the absence of part attributes and also reduces the misclassification errors. In PCC, the missing values in all the incomplete patterns are imputed using prototype of each class center, and the edited pattern with each imputation is classified by a standard classifier (for complete pattern). With PCC, one obtains *c* pieces of classification results for each incomplete pattern in a *c* class problem, and the global fusion of the *c* results is given for the credal classification. Unfortunately, PCC classifier is computationally greedy and time-consuming, and the imputation of missing values based on class prototype is not so precise. In order to overcome the limitations of PCC, we propose a new credal classification method for incomplete pattern with adaptive imputation of missing values, and it can be called Credal Classification with Adaptive Imputation (CCAI) for short.

The pattern to classify usually consists of multiple attributes. Sometimes, the class of the pattern can be precisely determined using only a part (a subset) of the available attributes, and it implies that the other attributes are redundant and in fact unnecessary for the classification. So a new method of credal classification with adaptive imputation strategy (i.e. CCAI) for missing values is proposed. In CCAI, we attempt to classify the object only using the known attributes value at first. If a specific classification result is obtained, it very likely means that the missing values are not very necessary for the classification, and we directly take the decision on the class of the object based on this result. However, if the object cannot be clearly classified with the available information, it indicates that the missing information included in the missing attribute values is probably very crucial for making the classification. In this case, we present a sophisticated classification strategy for the edition of pattern based on the proper imputation of missing values.

K-nearest neighbors-based imputation method usually provides pretty good performances for the estimation of missing values, but its main drawback is the big computational burden. To reduce the computational burden, Self-Organizing Map (SOM) [38] is applied in each class, and the optimized weighting vectors are used to represent the corresponding class. Then, the *K* nearest weighting vectors of the object in each class are employed to estimate the missing values. For the classification of original incomplete pattern (without imputation of missing values) or the edited pattern (with imputation of missing values), we adopt the ensemble classifier approach. One can get the simple classification result according to each training class, and each classification result is represented by a simple basic belief assignment (BBA) including two focal elements (i.e. singleton class and ignorant class) only. The belief of the object belonging to each class is calculated based on the distance to the corresponding prototype, and the other belief is committed to the ignorant element. The fusion (ensemble) of these multiple BBA's is then used to determine the class of the object. If the object is directly classified using only the known values, Dempster–Shafer[1] (DS) fusion rule [16] is applied because of the simplicity of this rule and also because the BBA's to fuse are usually in low conflict. In this case, a specific result is obtained with DS rule. Otherwise, a new fusion rule inspired by Dubois and Prade (DP) rule [39] is used to classify the edited pattern with proper imputation of its missing values. Because the estimation of the missing values can be quite uncertain, it naturally induces an imprecise classification. So the partial conflicting beliefs will be kept and committed to the associated meta-classes in this new rule to reasonably reveal the potential imprecision of the classification result.

[1] Although the rule has been proposed originally by Arthur Dempster, we prefer to call it Dempster–Shafer rule because it has been widely promoted by Shafer in [16].

In this paper, we present a credal classification method with adaptive imputation of missing values based on belief function theory for dealing with the incomplete patterns, and it is organized as follows. The basics of belief function theory and Self-Organizing Map is briefly recalled in Section 2. The new credal classification method for incomplete patterns is presented in the Section 3, and the proposed method is then tested and evaluated in Section 4 compared with several other classical methods. The paper is concluded in the final section.

2. Background knowledge

Belief function theory (BFT) can well characterize the uncertain and imprecise information, and it is used in this work for the classification of patterns. SOM technique is employed to find the optimized weighting vectors which are used to represent the corresponding class, and this can reduce the computation burden in the estimation of the missing values based on K-NN method. So the basic knowledge on BFT and SOM will be briefly recalled.

2.1. Basis of belief function theory

The Belief Function Theory (BFT) introduced by Glenn Shafer is also known as Dempster–Shafer Theory (DST), or the Mathematical Theory of Evidence [16–18]. Let us consider a frame of discernment consisting of c exclusive and exhaustive hypotheses (classes) denoted by $\Omega=\{\omega_i, i=1,2,\ldots,c\}$. The power-set of Ω denoted 2^{Ω} is the set of all the subsets of Ω, empty set included. For example, if $\Omega=\{\omega_1,\omega_2,\omega_3\}$, then $2^{\Omega}=\{\varnothing,\omega_1,\omega_2,\omega_3,\omega_1\cup\omega_2,\omega_1\cup\omega_3,\omega_2\cup\omega_3,\Omega\}$. In the classification problem, the singleton element (e.g. ω_i) represents a specific class. In this work, the disjunction (union) of several singleton elements is called a *meta-class* which characterizes the partial ignorance of classification. Examples of meta-classes are $\omega_i\cup\omega_j$, or $\omega_i\cup\omega_j\cup\omega_k$. In BFT, one object can be associated with different singleton elements as well as with sets of elements according to a basic belief assignment (BBA), which is a function $m(\cdot)$ from 2^{Ω} to $[0,1]$ satisfying $m(\varnothing)=0$ and the normalization condition $\sum_{A\in 2^{\Omega}} m(A)=1$. The subsets A of Ω such that $m(A)>0$ are called the focal elements of the belief mass $m(\cdot)$.

The credal classification (or partitioning) [29] is defined as n-tuple $M=(\mathbf{m_1},\ldots,\mathbf{m_n})$ of BBA's, 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 in the power-set 2^{Θ}. The credal classification allows the objects to belong to the specific classes and the sets of classes corresponding to meta-classes with different belief mass assignments. The credal classification can well model the imprecise and uncertain information thanks to the introduction of meta-class.

For combining multiple sources of evidence represented by a set of BBA's, the well-known Dempster's rule [16] is still widely used, even if its justification is an open debate and questionable in the community [40,41]. The combination of two BBA's $m_1(\cdot)$ and $m_2(\cdot)$ over 2^{Ω} is done with DS rule of combination defined by $m_{DS}(\varnothing)=0$ and for $A\neq\varnothing, B,C\in 2^{\Omega}$ by

$$m_{DS}(A)=\frac{\sum_{B\cap C=A} m_1(B)m_2(C)}{1-\sum_{B\cap C=\varnothing} m_1(B)m_2(C)} \tag{1}$$

DS rule is commutative and associative, and makes a compromise between the specificity and complexity for the combination of BBA's. With this rule, all the conflicting beliefs $\sum_{B\cap C=\varnothing} m_1(B)m_2(C)$ are proportionally redistributed back to the focal elements through a classical normalization step. However, this redistribution can yield unreasonable results in the high conflicting cases [40], as well as in some special low conflicting cases [41]. That is why different rules of combination have emerged to overcome its limitations. Among the possible alternatives of DS rule, we find Smets' conjunctive rule (used in his transferable belief model (TBM) [18]), Dubois–Prade (DP) rule [39], and more recently the more complex Proportional Conflict Redistributions (PCR) rules [42]. Unfortunately, DP and PCR rules are less appealing from implementation standpoint since they are not associative, and they become complex to use when more than two BBA's have to be combined altogether.

2.2. Overview of Self-Organizing Map

Self-Organizing Map (SOM) (also called Kohonen map) [38] introduced by Teuvo Kohonen is a type of artificial neural network (ANN), and it is trained by unsupervised learning method. SOM defines a mapping from the input space to a low-dimensional (typically two-dimensional) grid of $M\times N$ nodes. So it allows us to approximate the feature space dimension (e.g. a real input vector $\mathbf{x}\in\mathbb{R}^p$) into a projected 2D space, and it is still able to preserve the topological properties of the input space using a neighborhood function. Thus, SOM is very useful for visualizing low-dimensional views of high-dimensional data by a nonlinear projection.

The node at position $(i,j), i=1,\ldots M, j=1,\ldots,N$ corresponds to a weighting vector denoted by $\sigma(i,j)\in\mathbb{R}^p$. An input vector $\mathbf{x}\in\mathbb{R}^p$ is to be compared to each $\sigma(i,j)$, and the neuron whose weighting vector is the most close (similar) to $\mathbf{x}$ according to a given metric is called the best matching unit (BMU), which is defined as the output of SOM with respect to $\mathbf{x}$. In real applications, the Euclidean distance is usually used to compare $\mathbf{x}$ and $\sigma(i,j)$. The input pattern $\mathbf{x}$ can be mapped onto the SOM at location (i,j) where $\sigma(i,j)$ is with the minimal distance to $\mathbf{x}$. It is considered that the SOM achieves a non-uniform quantization that transforms $\mathbf{x}$ to $\sigma_{\mathbf{x}}$ by minimizing the given metric (e.g. distance measure) [43].

In SOM, the competitive learning is adopted, and the training algorithm is iterative. The initial values of the weighting vectors σ may be set randomly, but they will converge to a stable value at the end of the training process. When an input vector is fed to the network, its Euclidean distance to all weight vectors is computed. Then the BMU whose weight vector is most similar to the input vector is found, and the weights of the BMU and neurons close to it in the SOM grid are adjusted towards the input vector. The magnitude of the change decreases with time and with distance (within the grid) from the BMU. The detailed information about SOM can be found in [38].

In this work, SOM is applied in each training class to obtain the optimized weighting vectors that are used to represent the corresponding class. The number of the weighting vectors is much smaller than the original samples in the associated training class. We will utilize these weighting vectors rather than the original samples to estimate the missing values in the object (incomplete pattern), and this could effectively reduce the computation burden.

3. Credal classification of incomplete pattern

Our new method consists of two main steps. In the first step, the object (incomplete pattern) is directly classified according to the known attribute values only, and the missing values are ignored. If one can get a specific classification result, the classification procedure is done because the available attribute information is sufficient for making the classification. But if the class of the object cannot be clearly identified in the first step, it means that the unavailable information included in the missing values is likely crucial for the classification. In this case, one has to enter in the second step of the method to classify the object with a proper

imputation of missing values. In the classification procedure, the original or edited pattern will be classified according to each class of training data. The global fusion of these classification results, which can be considered as multiple sources of evidence represented by BBA's, is then used for the credal classification of the object. Our new method for credal classification of incomplete pattern with adaptive imputation of missing values is referred as Credal Classification with Adaptive Imputation, or just as CCAI for conciseness. CCAI is based on belief function theory, which can well manage the uncertain and imprecise information caused by the missing values in the classification.

3.1. First step: direct classification of incomplete pattern using the available data

Let us consider a set of test patterns (samples) $X=\{\mathbf{x}_1,\ldots,\mathbf{x}_n\}$ to be classified based on a set of labeled training patterns $Y=\{\mathbf{y}_1,\ldots,\mathbf{y}_s\}$ over the frame of discernment $\Omega=\{\omega_1,\ldots,\omega_c\}$. In this work, we focus on the classification of incomplete pattern in which some attribute values are absent. So we consider all the test patterns (e.g. $\mathbf{x}_i, i=1,\ldots,n$) with several missing values. The training data set Y may also have incomplete patterns in some applications. However, if the incomplete patterns take a very small amount say less than 5% in the training data set, they can be ignored in the classification. If the percentage of incomplete patterns is big, the missing values must usually be estimated at first, and the classifier will be trained using the edited (complete) patterns. In the real applications, one can also just choose the complete labeled patterns to include in the training data set when the training information is sufficient. So for simplicity and convenience, we consider that the labeled samples (e.g. $\mathbf{y}_j, j=1,\ldots,s$) of the training set Y are all complete patterns in the sequel.

In the first step of classification, the incomplete pattern say $\mathbf{x}_i$ will be classified according to each training class by a normal classifier (for dealing with the complete pattern) at first, and all the missing values are ignored here. In this work, we adopt a very simple classification method[2] for the convenience of computation, and $\mathbf{x}_i$ is directly classified based on the distance to the prototype of each class.

The prototype of each class $\{\mathbf{o}_1,\ldots,\mathbf{o}_c\}$ corresponding to $\{\omega_1,\ldots,\omega_c\}$ is given by the arithmetic average vector of the training patterns in the same class. Mathematically, the prototype is computed for $g=1,\ldots,c$ by

$$\mathbf{o}_g=\frac{1}{N_g}\sum_{\mathbf{y}_j\in\omega_g}\mathbf{y}_j \tag{2}$$

where N_g is the number of the training samples in the class ω_g.

In a c-class problem, one can get c pieces of simple classification result for $\mathbf{x}_i$ according to each class of training data, and each result is represented by a simple BBA's including two focal elements, i.e. the singleton class and the ignorant class (Ω) to characterize the full ignorance. The belief of $\mathbf{x}_i$ belonging to class ω_g is computed based on the distance between $\mathbf{x}_i$ and the corresponding prototype $\mathbf{o}_g$. Normalized Euclidean distance as Eq. (4) is adopted here to deal with the anisotropic class, and the missing values are ignored in the calculation of this distance. The other mass of belief is assigned to the ignorant class Ω. Therefore, the BBA's construction is done by

$$\begin{cases} m_i^{\mathbf{o}_g}(\omega_g)=e^{-\eta d_{ig}} \\ m_i^{\mathbf{o}_g}(\Omega)=1-e^{-\eta d_{ig}} \end{cases} \tag{3}$$

with

$$d_{ig}=\sqrt{\frac{1}{p}\sum_{j=1}^{p}\left(\frac{x_{ij}-o_{gj}}{\delta_{gj}}\right)^2} \tag{4}$$

and

$$\delta_{gj}=\sqrt{\frac{1}{N_g}\sum_{\mathbf{y}_i\in\omega_g}(y_{ij}-o_{gj})^2} \tag{5}$$

where x_{ij} is value of $\mathbf{x}_i$ in j-th dimension, and y_{ij} is value of $\mathbf{y}_i$ in j-th dimension. p is the number of available attribute values in the object $\mathbf{x}_i$. The coefficient $1/p$ is necessary to normalize the distance value because each test sample can have a different number of missing values. δ_{gj} is the average distance of all training samples in class ω_g to the prototype o_g in j-th dimension. N_g is the number of training samples in ω_g. η is a tuning parameter, and the bigger η generally yields smaller mass of belief on the specific class w_g. It is usually recommended to take $\eta\in[0.5,0.8]$ according to our various tests, and $\eta=0.7$ can be considered as default value.

Obviously, the smaller the distance measure, the bigger the mass of belief on the singleton class. This particular structure of BBA's indicates that we can just confirm the degree of the object $\mathbf{x}_i$ associated with the specific class ω_g only according to training data in ω_g. The other mass of belief reflects the level of belief one has on full ignorance, and it is committed to the ignorant class Ω. Similarly, one calculates c independent BBA's $m_i^{\mathbf{o}_g}(\omega_g), g=1,\ldots,c$ based on the different training classes.

Before combining these c BBA's, we examine whether a specific classification result can be derived from these c BBA's. This is done as follows: if it holds that $m_i^{\mathbf{o}_{1st}}(\omega_{1st})=\text{argmax}_g(m_i^{\mathbf{o}_g}(\omega_g))$, then the object will be considered to belong very likely to the class ω_{1st}, which obtains the biggest mass of belief in the c BBA's. The class with the second biggest mass of belief is denoted ω_{2nd}.

The distinguishability degree $\chi_i\in(0,1]$ of an object $\mathbf{x}_i$ associated with different classes is defined by

$$\chi_i=\frac{m_i^{\mathbf{o}_{2nd}}(\omega_{2nd})}{m_i^{\mathbf{o}_{max}}(\omega_{max})} \tag{6}$$

Let ϵ be a chosen small positive distinguishability threshold value in $(0,1]$. If the condition $\chi_i\le\epsilon$ is satisfied, it means that all the classes involved in the computation of χ_i can be clearly distinguished of $\mathbf{x}_i$. In this case, it is very likely to obtain a specific classification result from the fusion of the c BBA's. The condition $\chi_i\le\epsilon$ also indicates that the available attribute information is sufficient for making the classification of the object, and the imputation of the missing values is not necessary. If $\chi_i\le\epsilon$ condition holds, the c BBA's are directly combined with DS rule to obtain the final classification results of the object because DS rule usually produces specific combination result with acceptable computation burden in the low conflicting case. In such case, the meta-class is not included in the fusion result, because these different classes are considered distinguishable based on the condition of distinguishability. Moreover, the mass of belief of the full ignorance class Ω, which represents the noisy data (outliers), can be proportionally redistributed to other singleton classes for more specific results if one knows a priori that the noisy data is not involved.

If the distinguishability condition $\chi_i\le\epsilon$ is not satisfied, it means that the classes ω_{1st} and ω_{2nd} cannot be clearly distinguished for the object with respect to the chosen threshold value ϵ, indicating that missing attribute values play almost surely a crucial role in the classification. In this case, the missing values must be properly imputed to recover the unavailable attribute information before entering the classification procedure. This is the Step 2 of our method which is explained in the next subsection.

[2] Many other normal classifiers (e.g. K-NN) can be selected here depending on the preference of user, and we propose to use this simple classification method because of its low computation complexity.

3.2. Second step: classification of incomplete pattern with imputation of missing values

3.2.1. Multiple estimation of missing values

In the estimation of the missing attribute values, there exist various methods. Particularly, the K-NN imputation method generally provides good performance. However, the main drawback of KNN method is its big computational burden, since one needs to calculate the distances of the object with all the training samples. Inspired by [43], we propose to use the Self-Organized Map (SOM) technique [38] to reduce the computational complexity. SOM can be applied in each class of training data, and then $M \times N$ weighting vectors will be obtained after the optimization procedure. These optimized weighting vectors allow us to characterize well the topological features of the whole class, and they will be used to represent the corresponding data class. The number of the weighting vectors is usually small (e.g. 5×6). So the K nearest neighbors of the test pattern associated with these weighting vectors in the SOM can be easily found with low computational complexity.[3] The selected weighting vector no. k in the class ω_g, $g = 1, \ldots, c$, is denoted $\sigma_k^{\omega_g}$, for $k = 1, \ldots, K$.

In each class, the K selected close weighting vectors provide different contributions (weight) in the estimation of missing values, and the weight $p_{ik}^{\omega_g}$ of each vector is defined based on the distance between the object $\mathbf{x}_i$ and weighting vector $\sigma_k^{\omega_g}$:

$$p_{ik}^{\omega_g} = e^{(-\lambda d_{ik}^{\omega_g})} \tag{7}$$

with

$$\lambda = \frac{cNM(cNM-1)}{2\sum_{i,j} d(\sigma_i, \sigma_j)} \tag{8}$$

where $d_{ik}^{\omega_g}$ is the Euclidean distance between $\mathbf{x}_i$ and the neighbor $\mathbf{o}_k^{\omega_g}$ ignoring the missing values, and $1/\lambda$ is the average distance between each pair of weighting vectors produced by SOM in all the classes; c is the number of classes; $M \times N$ is the number of weighting vectors obtained by SOM in each class; and $d(\sigma_i, \sigma_j)$ is the Euclidean distance between any two weighting vectors σ_i and σ_j.

The weighted mean value $\hat{\mathbf{y}}_i^{\omega_g}$ of the selected K weighting vectors in training class ω_g will be used for the imputation of missing values. It is calculated by

$$\hat{\mathbf{y}}_i^{\omega_g} = \frac{\sum_{k=1}^{K} p_{ik}^{\omega_g} \sigma_k^{\omega_g}}{\sum_{k=1}^{K} p_{ik}^{\omega_g}} \tag{9}$$

The missing values in $\mathbf{x}_i$ will be filled by the values of $\hat{\mathbf{y}}_i^{\omega_g}$ in the same dimensions. By doing this, we get the edited pattern $\mathbf{x}_i^{\omega_g}$ according to the training class ω_g.

Then $\mathbf{x}_i^{\omega_g}$ will be simply classified only based on the training data in ω_g as similarly done in the direct classification of incomplete pattern using Eq. (3) of Step 1 for convenience.[4]

The classification of $\mathbf{x}_i$ with the estimation of missing values is also done based on the other training classes according to this procedure. For a c-class problem, there are c training classes, and therefore one can get c pieces of classification results with respect to one object.

[3] The training of SOM using the labeled patterns becomes time consuming when the number of labeled patterns is big, but fortunately it can be done off-line. In our experiments, the running time performance shown in the results does not include the computational time spent for the off-line procedures.

[4] Of course, some other sophisticated classifiers can also be applied here according to the selection of user, but the choice of classifier is not the main purpose of this work.

3.2.2. Ensemble classifier for credal classification

These c pieces of results obtained by each class of training data in a c-class problem are considered with different weights, since the estimations of the missing values according to different classes have different reliabilities. The weighting factor of the classification result associated with the class w_g can be defined by the sum of the weights of the K selected SOM weighting vectors for the contributions to the missing values imputation in ω_g, which is given by

$$\rho_i^{\omega_g} = \sum_{k=1}^{K} p_{ik}^{\omega_g} \tag{10}$$

The result with the biggest weighting factor $\rho_i^{\omega_{max}}$ is considered as the most reliable, because one assumes that the object must belong to one of the labeled classes (i.e. w_g, $g = 1, \ldots, c$). So the biggest weighting factor will be normalized as one. The other relative weighting factors are defined by

$$\hat{\alpha}_i^{\omega_g} = \frac{\rho_i^{\omega_g}}{\rho_i^{\omega_{max}}} \tag{11}$$

If the condition[5] $\hat{\alpha}_i^{\omega_g} < \epsilon$ is satisfied, the corresponding estimation of the missing values and the classification result are not very reliable. Very likely, the object does not belong to this class. It is implicitly assumed that the object can belong to only one class in reality. If this result whose relative weighting factor is very small (w.r.t. ϵ) is still considered useful, it will be (more or less) harmful for the final classification of the object. So if the condition $\hat{\alpha}_i^{w_g} < \epsilon$ holds, then the relative weighting factor is set to zero. More precisely, we will take

$$\alpha_i^{\omega_g} = \begin{cases} 0 & \text{if } \hat{\alpha}_i^{\omega_g} < \epsilon \\ \dfrac{\rho_i^{\omega_g}}{\rho_i^{\omega_{max}}} & \text{otherwise.} \end{cases} \tag{12}$$

After the estimation of weighting (discounting) factors $\alpha_i^{\omega_g}$, the c classification results (the BBA's $m_i^{\mathbf{o}_g}(\cdot)$) are classically discounted [16] by

$$\begin{cases} \hat{m}_i^{\mathbf{o}_g}(\omega_g) = \alpha_i^{\omega_g} m_i^{\mathbf{o}_g}(\omega_g) \\ \hat{m}_i^{\mathbf{o}_g}(\Omega) = 1 - \alpha_i^{\omega_g} + \alpha_i^{\omega_g} m_i^{\mathbf{o}_g}(\Omega) \end{cases} \tag{13}$$

These discounted BBA's will be globally combined to get the credal classification result. If $\alpha_i^{\omega_g} = 0$, one gets $\hat{m}_i^{\mathbf{o}_g}(\Omega) = 1$, and this fully ignorant (vacuous) BBA plays a neutral role in the global fusion process for the final classification of the object.

Although we have done our best to estimate the missing values, the estimation can be quite imprecise when the estimations are obtained from different classes with the similar weighting factors, and the different estimations probably lead to distinct classification results. In such case, we prefer to cautiously keep (rather to ignore) the uncertainty, and maintain the uncertainty in the classification result. Such uncertainty can be well reflected by the conflict of these classification results represented by the BBA's. DS rule is not suitable here, because all the conflicting beliefs are distributed to other focal elements. A particular combination rule inspired by DP rule is introduced here to fuse these BBA's according to the current context. In our new rule, the partial conflicting beliefs are prudently transferred to the proper meta-class to reveal the imprecision degree of the classification caused

[5] The threshold ϵ is the same as in Section 3.1, because it is also used to measure the distinguishability degree here.

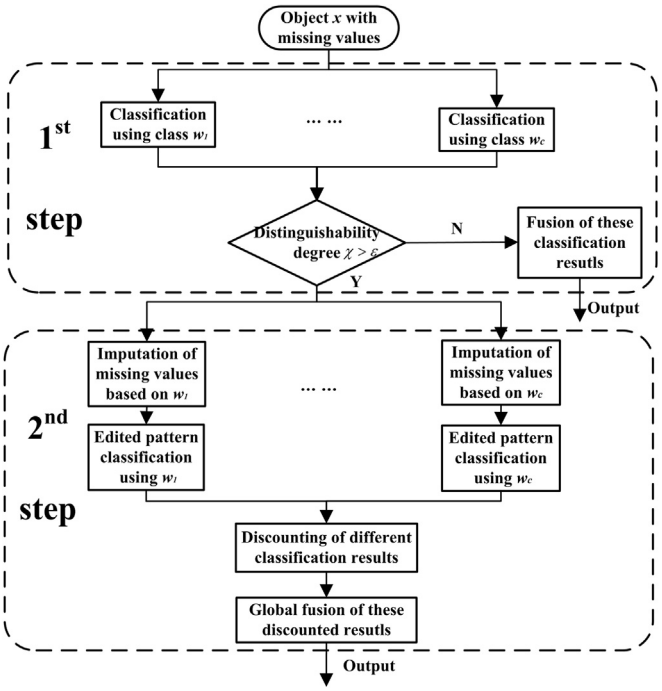

Fig. 1. Flowchart of the proposed CCAI method.

by the missing values. This new rule of combination is defined by

$$\begin{cases} m_i(\omega_g) = \hat{m}_i^{\mathbf{o}_g}(\omega_g) \prod_{j \neq g} \hat{m}_i^{\mathbf{o}_j}(\Omega) \\ m_i(A) = \prod_{\bigcup_j \omega_j = A} \hat{m}_i^{\mathbf{o}_j}(\omega_j) \prod_{k \neq j} \hat{m}_i^{\mathbf{o}_k}(\Omega) \end{cases} \tag{14}$$

The test pattern can be classified according to the fusion results, and the object is considered belonging to the class (singleton class or meta-class) with the maximum mass of belief. This is called hard credal classification. If one object is classified into a particular class, it means that this object has been correctly classified with the proper imputation of missing values. If one object is committed to a meta-class (e.g. $A \cup B$), it means that we just know that this object belongs to one of the specific classes (e.g. A or B) included in the meta-class, but we cannot specify which one. This case can happen when the missing values are essential for the accurate classification of this object, but the missing values cannot be estimated very well according to the context, and different estimations will induce the classification of the object into distinct classes (e.g. A or B).

For convenience, Fig. 1 shows the functional flowchart of this new CCAI method.

Guideline for tuning of the parameters ϵ and η: The tuning of parameters η and ϵ is very important in the application of CCAI. η in Eq. (3) is associated with the calculation of mass of belief on the specific class, and the bigger η value will lead to smaller mass of belief committed to the specific class. Based on our various tests, we advise to take $\eta \in [0.5, 0.8]$, and the value $\eta = 0.7$ can be taken as the default value. The parameter ϵ is the threshold to tune for changing the classification strategy. It is also used in Eq. (12) for the calculation of the discounting factor. The bigger ϵ will make fewer objects going to the sophisticated classification procedure with the imputation of missing values, and it also forces more discounting factors to zero according to Eq. (12), which implies that fewer simple classification results obtained based on each class can be useful in the global fusion step. So the bigger ϵ will make fewer objects committed to the meta-classes (corresponding to the low imprecision of classification), but it increases the risk of misclassification error. ϵ should be tuned according to the compromise one can accept between the misclassification error and imprecision (non specificity of classification decision). One can also apply the cross validation [44] (e.g. leave-one-out method) in the training data space to find a suitable threshold, and the missing values in the test samples are randomly distributed in all the dimensions.

4. Experiments

Three experiments with artificial and real data sets have been used to test the performance of this new CCAI method compared with the K-NN imputation (KNNI) method [12], FCM imputation (FCMI) method [13,14], SOM imputation (SOMI) [15] method and our previous credal classification PCC method [25]. SOM technique is also employed in the second step of CCAI method, but CCAI is different from the previous SOMI method. In SOMI method, SOM is applied for the whole training data set, and the missing values are precisely estimated based on an activation group composed of the best match node (unit) of input pattern and its close neighbors. Then, the edited pattern with the imputation of missing values can be classified using a standard classifier. Nevertheless, SOM is not involved in the first step of CCAI, and the object is directly classified ignoring the missing values. In the second step of CCAI, SOM is applied in each training class, and multiple estimations of missing values can be obtained based on the input pattern's K nearest weighting vectors corresponding to nodes of SOM in each class. Then different classification results will be produced according to different estimations, and these results are globally fused for final classification. The conflicting information committed to the meta-class is kept in the fusion to characterize the imprecision of classification in CCAI, but this cannot be done in SOMI. These different methods have been programmed and tested with Matlab™ software.

The evidential neural network classifier (ENN) [27] is adopted in the sequel experiments to classify the edited pattern with the estimated values in PCC, KNNI and FCMI, since ENN produces generally good results in the classification.[6] The evidential K-nearest neighbor (EK-NN) method [21] is also used to classify the edited pattern in Experiment 3 with real data for comparison. The parameters of ENN and EK-NN can be automatically optimized as explained in [27] and [22] respectively. In SOMI, we use the $M \times N = 6 \times 8$ nodes for mapping the whole input data set consisting of all the training classes to the 2-dimensional grid, and it has good performance. In the applications of PCC, the tuning parameter ϵ can be tuned according to the imprecision rate one can accept. In CCAI, a small number of the nodes in the 2-dimensional grid of SOM are given by $M \times N = 3 \times 4$ for each class, and we take the value of $K = N = 4$ in K-NN for the imputation of missing values. This seems to provide good result in the sequel experiments. In order to show the ability of CCAI and PCC to deal with the meta-classes, the hard credal classification is applied, and the class of each object is decided according to the criterion of the maximal mass of belief.

In our simulations, the misclassification is declared (counted) for one object truly originated from ω_i if it is classified into A with $\omega_i \cap A = \varnothing$. If $\omega_i \cap A \neq \varnothing$ and $A \neq \omega_i$ then it will be considered as an imprecise classification. The error rate denoted by Re is calculated by $Re = N_e/T$, where N_e is number of misclassification errors, and T is the number of objects under test. The imprecision rate denoted by Ri_j is calculated by $Ri_j = Ni_j/T$, where Ni_j is number of

[6] Other traditional classifiers for complete pattern can also be selected here according to the actual application.

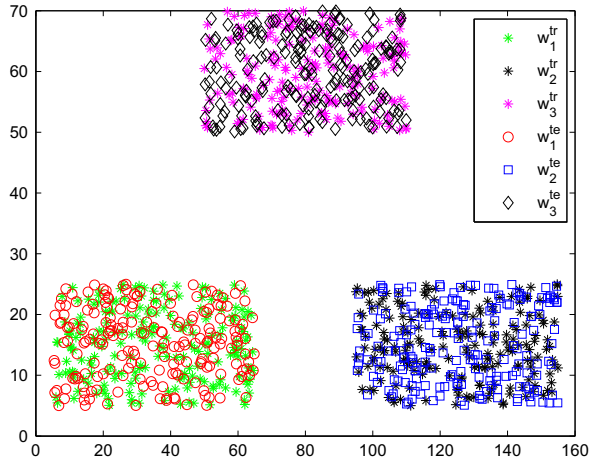

Fig. 2. Training data and test data.

objects committed to the meta-classes with the cardinality value j. In our experiments, the classification of object is generally uncertain (imprecise) among a very small number (e.g. 2) of classes, and we only take Ri_2 here since there is no object committed to the meta-class including three or more specific classes.

4.1. Experiment 1 (artificial data set)

In the first experiment, we show the interest of credal classification based on belief functions with respect to the traditional classification working with probability framework. A 3-class data set $\Omega=\{\omega_1,\omega_2,\omega_3\}$ obtained from three 2-D uniform distributions shown by Fig. 2, is considered here. Each class has 200 training samples and 200 test samples, and there are 600 training samples and 600 test samples in total. The uniform distributions of the three classes are characterized by the following interval bounds:

	x-label interval	y-label interval
ω_1	(5, 65)	(5, 25)
ω_2	(95, 155)	(5, 25)
ω_3	(50, 110)	(50, 70)

The values in the second dimension corresponding to the y-coordinate of test samples are all missing. So test samples are classified according to the only one available value in the first dimension corresponding to x-coordinate.

Several different methods like FCMI, KNNI, SOMI have been applied here for comparison with CCAI as shown by Fig. 3(a)–(f). Particularly, the classification result obtained using the (first or second) single step of CCAI (denoted by SCCAI) is also given as in Fig. 3(d)–(e). In the first step of CCAI, the direct classification is done without imputation of missing value, whereas the object is classified with imputation of missing values in all incomplete patterns by only the second step of CCAI.

A particular value of $K=9$ is selected in the classifier K-NN imputation method.[7] For notation conciseness, we have denoted $\omega^{te}\triangleq\omega^{test}$, $\omega^{tr}\triangleq\omega^{training}$ and $\omega_{i,\ldots,k}\triangleq\omega_i\cup\ldots\cup\omega_k$. The error rate (in %), imprecision rate (in %) and computation time (s) are specified in the caption of each subfigure.

Because the y value in the test sample is missing, the class w_3 appears partially overlapped with the classes ω_1 and ω_2 on their margins according to the value of the x-coordinate as shown in Fig. 3(a). The missing value of the samples in the overlapped parts can be filled by quite different estimations obtained from different classes with the almost same reliabilities. For example, the estimation of the missing values of the objects in the right margin of ω_1 and the left margin of ω_3 can be obtained according to the training class ω_1 or ω_3. The edited pattern with the estimation from ω_1 will be classified into class ω_1, whereas it will be committed to class ω_3 if the estimation is drawn from ω_3. It is similar to the test samples in the left margin of ω_2 and the right margin of ω_3. This indicates that the missing value plays a crucial rule in the classification of these objects, but unfortunately the estimation of these involved missing values are quite uncertain according to context. So these objects are prudently classified into the proper meta-class (e.g. $\omega_1\cup\omega_3$ and $\omega_2\cup\omega_3$) by CCAI. The CCAI results indicate that these objects belong to one of the specific classes included in the meta-classes, but these specific classes cannot be clearly distinguished by the object based only on the available values. If one wants to get more precise and accurate classification results, one needs to request for additional resources for gathering more useful information. The other objects in the left margin of ω_1, right margin of ω_2 and middle of ω_3 can be correctly classified based on the only known value in the x-coordinate, and it is not necessary to estimate the missing value for the classification of these objects in CCAI. However, all the test samples are classified into specific classes by the traditional methods KNNI and FCMI, and this causes many errors due to the limitation of probability framework. If we just apply the first step of SCCAI without imputation of the missing value and directly classify all the objects using the only known value (i.e. value in the x-coordinate), it produces bigger error rate than the other methods, and this indicates that the imputation procedure is important to improve the accuracy of classification. If only the second step of SCCAI is done with imputation of the missing values in all incomplete patterns, it causes high imprecision rate that is not an efficient solution, and it takes much longer computation time than CCAI. CCAI with the adaptive imputation strategy can well balance the error rate, imprecision rate and computation burden. CCAI consisting of two steps generally produces smaller error rate than KNNI, FCMI and SOMI thanks to the use of meta-classes. Meanwhile, the computational time of CCAI is similar to that of FCMI, and is much shorter than KNNI because of the introduction of SOM technique in the estimation of missing values. It shows that the computational complexity of CCAI is relatively low. This simple example shows the interest and the potential of the credal classification obtained with CCAI method.

4.2. Experiment 2 (artificial data set)

In this second experiment, we evaluate the performance of CCAI method using a 4D data set which includes 3 classes ω_1, ω_2, and ω_3. The artificial data are generated from three 4D Gaussian distributions characterized by the following means, vectors and covariance matrices ($\mathbf{I}$ denotes the 4×4 identity matrix):

$$\mu_1=(10,50,100,100)^T,\quad \mathbf{\Sigma}_1=10\cdot\mathbf{I}$$
$$\mu_2=(30,40,50,90)^T,\quad \mathbf{\Sigma}_2=15\cdot\mathbf{I}$$
$$\mu_3=(20,80,90,130)^T,\quad \mathbf{\Sigma}_3=12\cdot\mathbf{I}$$

We have used g training samples, and g test samples (for $g=500$, and $g=1000$) in each class. So there are total $N=3\times g$ training samples and $N=3\times g$ test samples. Each test sample has n missing values (for $n=1,2,3$), and the missing component value is randomly distributed in every dimension. Three other methods KNNI, FCMI, SOMI and PCC are also applied here for the

[7] In fact, the choice of K ranking from 7 to 15 does not affect seriously the results.

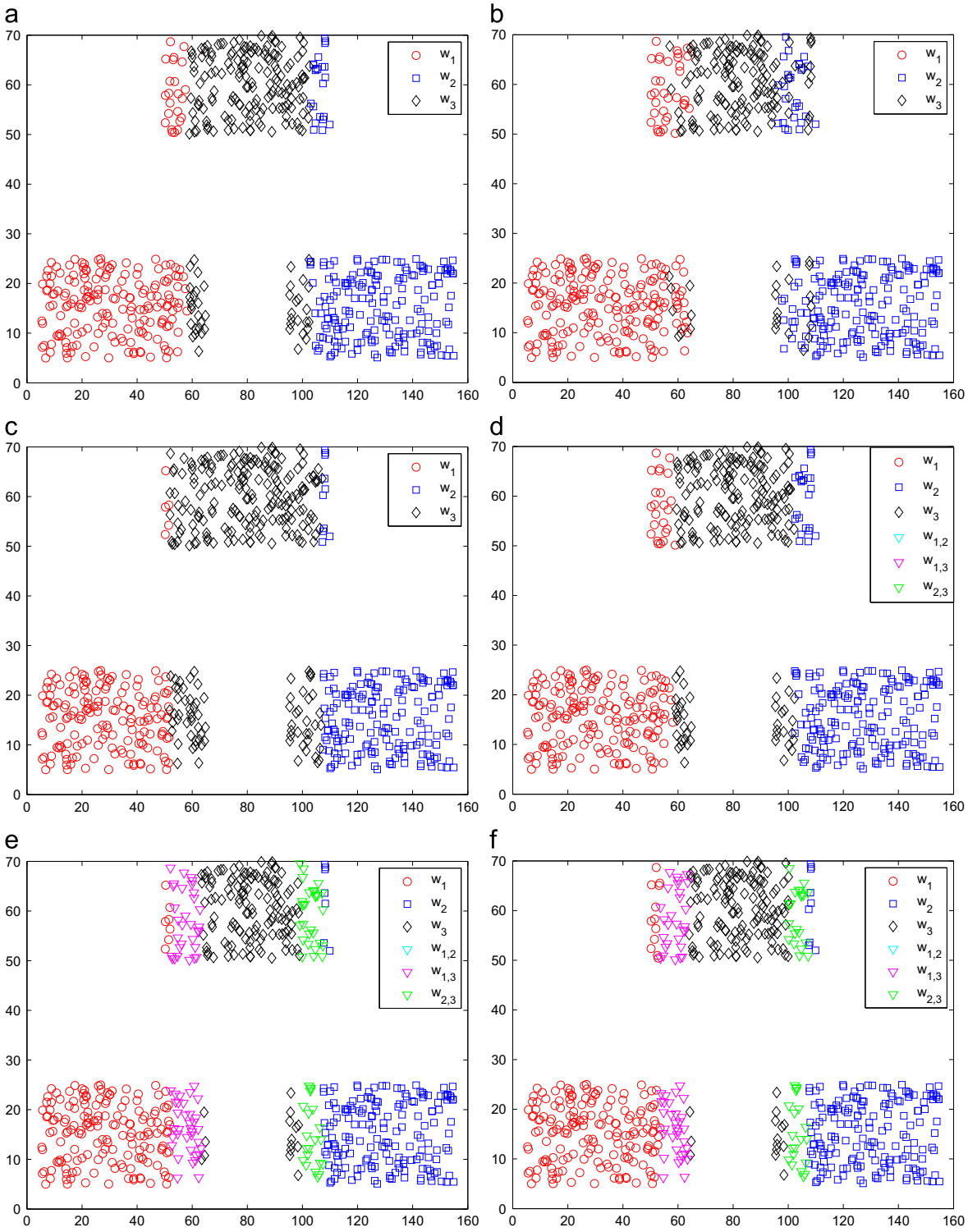

Fig. 3. Classification results of a 3-class artificial data set by different methods. (a). Classification result by FCMI ($Re = 14.67, time = 0.0469$ s). (b). Classification result by KNNI ($Re = 14.17, time = 7.9531$ s). (c). Classification result by SOMI ($Re = 14.33, time = 0.9063$ s). (d). Classification result only by 1st step of SCCAI ($Re = 14.83, time = 0.0156$ s). (e). Classification result only by 2nd step of SCCAI ($Re = 4.83, Ri_2 = 19.33, time = 0.1719$ s). (f). Classification result by CCAI ($Re = 5.83, Ri_2 = 16.83, time = 0.0469$ s).

performances comparison. For each pair (N, n), the reported error rates, imprecision rates and running time (s) are the averages over 10 trials performed with 10 independent random generation of the data sets. For KNNI, the values of K ranging from 5 to 20 neighbors have been tested, and the mean error rate with $K \in [5, 20]$ is given in Table 1. In PCC method, the parameter ϵ has been optimized to obtain an acceptable compromise between error rate and the imprecision degree. ENN is adopted to classify the edited pattern with imputation of missing values in FCMI, KNNI, SOMI and PCC.

The classification results of the applied methods (i.e. FCMI, KNNI, SOMI, PCC and CCAI) have been shown in Table 1. Our proposed CCAI method produces the lowest error rate, since some objects hard to correctly classify because of the missing values have been committed to the proper meta-class. Meanwhile, CCAI

Table 1
Classification results for 3-class data set by different methods (in %).

(N, n)	FCMI $\{Re, time\}$	KNNI $\{Re, time\}$	SOMI $\{Re, time\}$	PCC $\{Re, Ri_2, time\}$	CCAI $\{Re, Ri_2, time\}$
(1500,1)	{6.73, 0.9094 s}	{7.42, 3.0005 s}	{7.22, 0.9814 s}	{6.20, 2.33, 0.3484 s}	{4.64, 3.87, 0.2500 s}
(1500,2)	{14.38, 0.9016 s}	{15.68, 2.7759 s}	{15.43, 0.9546 s}	{13.47, 5.93, 0.3141 s}	{9.76, 9.79, 0.2344 s}
(1500,3)	{36.84, 0.9391 s}	{40.11, 3.002 s}	{40.10, 1.0322 s}	{34.57, 7.97, 0.3484 s}	{29.71, 15.6, 0.2906 s}
(3000,1)	{6.75, 1.3922 s}	{7.54, 12.0386 s}	{7.14, 1.7310 s}	{6.17, 1.63, 0.5453 s}	{4.73, 3.83, 0.3469 s}
(3000,2)	{14.73, 1.5375 s}	{15.80, 11.3857 s}	{15.20, 1.8203 s}	{14.00, 1.60, 0.5234 s}	{9.90, 10.33, 0.3063 s}
(3000,3)	{36.43,1.6500 s}	{40.48, 10.2803 s}	{40.05, 1.6094 s}	{33.94, 8.13, 0.5484 s}	{29.52, 16.83, 0.3937 s}

takes the shortest computation time compared with the other methods. This is because some incomplete patterns are directly classified ignoring the missing values, which are considered unimportant for the classification. However, the missing values in each pattern are all imputed by other methods, and this needs more computations and thus increases the computational time. Moreover, one can see that KNNI takes the longest time, and this is the main drawback of K-NN based method. The K-NN strategy is also adopted in CCAI, but we use a few optimized weighting vectors acquired by SOM technique to represent the whole training data class. Thus, we just need to calculate the distances between the object and these obtained weighting vectors rather than all the training samples, which reduces a lot the computation burden.

4.3. Experiment 3 (real data set)

Nine well known real data sets[8] available from UCI Machine Learning Repository [45] are used in this experiment to evaluate the performance of CCAI with respect to KNNI, FCMI, SOMI and PCC. Both ENN and EK-NN are employed here as standard classifier to classify the edited patterns. Moreover, the single (1st and 2nd) step procedure of CCAI (SCCAI) has been also applied here for comparison. In the first step of SCCAI, the object is directly classified using the only available attributes without imputation procedure, whereas all the missing values are imputed before the classification in the second step of SCCAI. The basic information of these used real data sets is given in Table 2. In Hepatitis data set, many patterns have already contained missing values. The patterns with missing values are considered as test samples, and the others are used as training samples. There is no missing values in the other seven original data sets, and it is assumed that n values are missing completely at random in all dimensions of each test sample. The cross validation is performed for these seven data sets, and we use the simplest 2-fold cross validation[9] 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 2-fold cross validation has been repeated 10 times, and the average error rate Re and imprecision rate Ri (for PCC and CCAI) of the different methods are given in Table 3. Particularly, the reported classification result of KNNI is the average with K value ranging from 5 to 15. For the notation conciseness, the selected classifier (SC) is denoted by A=EK-NN, B=ENN in Table 3. For the method of single step of CCAI (SCCAI), A represents the first step of SCCAI, and B represents the second step of SCCAI.

One can see in Table 3 that the credal classification of PCC and CCAI always produce the lower error rate than the traditional FCMI, KNNI and SOMI methods, since some objects that cannot be correctly classified using only the available attribute values have been properly committed to the meta-classes, which can well reveal the imprecision of classification. The selected classifiers (i.e. EK-NN and ENN) for classification of edited patterns in FCMI, KNNI, SOMI and PCC are usually with the similar performance in many cases,[10] but it is known that the K-NN based method generally has big computation burden. The choice of EK-NN and ENN should be made according to the actual condition in real applications. In CCAI, some objects with the imputation of missing values are still classified into the meta-class. It indicates that these missing values play a crucial role in the classification, but the estimation of these missing values is not very good. In other words, the missing values can be filled with the similar reliabilities by different estimated data, which lead to distinct classification results. So we have to cautiously assign them to the meta-class to reduce the risk of misclassification. Compared with our previous method PCC, this new method CCAI generally provides better performance with lower error rate and imprecision rate, and it is mainly because more accurate estimation method (i.e. SOM+KNN) for missing values is adopted in CCAI. However, if only the first step of SCCAI is applied, it produces more misclassification errors that other methods due to the absence of imputation of missing data. Whereas, the imprecision rate will be quite high if only the second step of SCCAI is adopted because all the conflicting beliefs caused in the combination procedure are transferred to the meta-classes. So CCAI with adaptive imputation of missing values can provide a good compromise between the error and imprecision. This third experiment using real data sets shows the effectiveness and interest of this new CCAI method with respect to other methods.

Table 2
Basic information of the used data sets.

Name	Classes	Attributes	Instances
Breast	2	9	699
Hepatitis	2	19	155
Statlog (Heart)	2	13	270
Iris	3	4	150
Seeds	3	7	210
Wine	3	13	178
Knowledge	4	5	403
Vehicle	4	18	946
Yeast	7	8	1429

5. Conclusion

A new credal classification method with adaptive imputation of missing values (called CCAI) for dealing with incomplete pattern has been presented based on belief function theory. In the first step of CCAI method, some objects (incomplete pattern) are

[8] We select seven classes from Yeast data set, because the last three classes (i.e. VAC POX and ERL) contain quite few samples.

[9] More precisely, the samples in each class 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.

[10] EK-NN outperforms ENN sometimes, but ENN can be better in some other cases.

Table 3
Classification results for different real data sets (rates in %).

Data set	(n, SC)	FCMI Re	KNNI Re	SOMI Re	PCC $\{Re, Ri_2\}$	SCCAI $\{Re, Ri_2\}$	CCAI $\{Re, Ri_2\}$
Hepatitis	A	26.40	27.38	27.47	{22.22, 7.56}	{23.67, 0}	{21.33, 5.33}
	B	25.33	26.67	25.33	{20.00, 6.67}	{20.00, 8.00}	
Breast	(3,A)	3.96	4.83	3.85	{4.39, 2.20}	{4.98, 0}	{3.66, 0}
	(3,B)	3.81	3.95	3.51	{3.81, 2.34}	{3.22, 0.73}	
	(6,A)	6.18	9.07	6.47	{5.82, 1.93}	{6.15, 0}	{4.83, 1.61}
	(6,B)	7.32	8.20	5.93	{5.42, 1.32}	{4.72, 2.93}	
	(7,A)	12.02	14.00	13.62	{10.11, 2.86}	{12.15, 0}	{9.00, 0.66}
	(7,B)	11.42	11.54	12.45	{10.10, 2.64}	{7.03, 17.11}	
Iris	(1,A)	6.89	5.29	5.14	{4.80, 2.04}	{6.67, 0}	{4.00, 1.33}
	(1,B)	7.33	4.89	5.00	{5.33, 2.67}	{4.00, 3.33}	
	(2,A)	13.89	13.02	13.24	{8.31, 6.27}	{12.00, 0}	{8.00, 4.67}
	(2,B)	14.00	11.33	12.67	{8.67, 4.00}	{7.33, 8.00}	
	(3,A)	18.22	18.67	18.00	{13.33, 8.67}	{17.33, 0}	{11.33, 12.00}
	(3,B)	17.33	18.44	17.34	{12.67, 9.33}	{10.67, 16.00}	
Seeds	(2,A)	15.56	11.59	11.63	{10.51, 2.95}	{9.52, 0}	{9.52, 0}
	(2,B)	15.24	11.19	10.20	{9.52, 4.76 }	{9.52, 0.95}	
	(4,A)	18.17	12.70	12.86	{10.22, 3.52}	{10.48, 0}	{10.00, 0.48}
	(4,B)	17.14	11.98	12.59	{10.48, 4.29}	{9.52, 1.90}	
	(6,A)	21.75	26.41	25.65	{17.84, 10.32}	{22.86, 0}	{16.19, 13.81}
	(6,B)	20.95	25.71	24.63	{16.19, 14.76}	{8.10, 28.57}	
Wine	(3,A)	29.32	27.12	27.53	{27.38, 0.71}	{6.97, 0}	{6.74, 1.12}
	(3,B)	26.97	26.97	28.65	{26.97, 1.69}	{6.18, 8.43}	
	(7,A)	34.68	26.22	31.30	{27.12, 0.79}	{7.87, 0}	{7.30, 3.93}
	(7,B)	33.24	30.43	31.46	{29.78, 2.25}	{5.62, 9.55}	
	(11,A)	34.76	29.55	34.35	{29.06, 1.61}	{14.61, 0}	{12.36, 3.93}
	(11,B)	33.43	30.90	32.58	{30.34, 2.81}	{10.67, 40.45}	
Knowledge	(1,A)	30.07	28.53	29.78	{26.72, 4.05}	{27.55, 0}	{ 20.85, 6.20}
	(1,B)	34.50	33.51	33.88	{28.35, 6.31}	{20.10, 8.19}	
	(2,A)	33.06	29.66	31.51	{27.32, 5.36}	{30.69, 0}	{23.57, 6.95}
	(2,B)	39.68	39.43	41.69	{33.32, 7.73}	{20.35, 13.40}	
	(3,A)	34.32	32.96	35.24	{29.86, 9.97}	{34.16, 0}	{30.51, 7.69}
	(3,B)	39.96	40.69	42.04	{33.76, 11.82}	{22.08, 21.59}	
Heart	(1,A)	37.41	37.78	36.67	{33.41, 12.59}	{17.78, 0}	{16.30, 0.37}
	(1,B)	41.18	41.85	41.11	{36.30, 9.63}	{13.70, 21.48}	
	(5,A)	48.15	38.27	41.48	{35.06, 25.93}	{23.70, 0}	{22.96, 0.74}
	(5,B)	46.89	43.09	42.96	{32.96, 28.52}	{22.59, 8.89}	
Vehicle	(5,A)	46.00	41.13	41.25	{35.63, 25.75}	{50.71, 0}	{34.87, 26.48}
	(5,B)	56.66	55.67	54.73	{37.87, 27.43}	{27.66, 50.24}	
	(9,A)	57.97	45.27	45.68	{38.63, 22.73}	{52.25, 0}	{36.64, 22.34}
	(9,B)	61.82	57.92	57.71	{43.63, 26.95}	{28.61, 56.97}	
Yeast	(1,A)	46.57	46.04	45.51	{42.71, 11.12}	{46.67, 0}	{40.28, 12.36}
	(1,B)	44.97	44.72	44.86	{39.86, 13.92}	{27.08, 46.74}	
	(3,A)	54.29	54.22	54.88	{51.86, 10.87}	{56.74, 0}	{49.75, 12.64}
	(3,B)	51.72	52.81	53.89	{49.38, 13.69}	{34.38, 49.31}	

directly classified ignoring the missing values if the specific classification result can be obtained, which effectively reduces the computation complexity because it avoids the imputation of the missing values. However, if the available information is not sufficient to achieve a specific classification of the object in the first step, we estimate (recover) the missing values before entering the classification procedure in a second step. The SOM and K-NN techniques are applied to make the estimation of missing attributes with a good compromise between the estimation accuracy and computation burden. The credal classification in this work allows the object to belong to different singleton classes and meta-class (i.e. disjunction of several classes) with different masses of belief. Once the object is committed to a meta-class (e.g. $A \cup B$), it means that the missing values cannot be accurately recovered according to the context, and the estimation is not very good. Different estimations will lead the object to distinct classes (e.g. A or B) involved in the meta-class. So some other sources of information will be required to achieve more precise classification of the object if necessary. The credal classification is able to well capture the imprecision of classification thanks to the meta-class and it effectively reduces the misclassification errors. The effectiveness and interest of the proposed CCAI method have been evaluated on three distinct experiments using artificial and real data sets.

Conflict of interest

None declared.

Acknowledgements

This work has been partially supported by National Natural Science Foundation of China – China (Nos. 61135001, 61403310), the Fundamental Research Funds for the Central Universities – China (No. 3102014JCQ01067) and the Natural fund of Shaanxi Province – China (2015JQ6265).

References

[1] P. Garcia-Laencina, J. Sancho-Gomez, A. Figueiras-Vidal, Pattern classification with missing data: a review, Neural Comput. Appl. 19 (2010) 263–282.
[2] R.J. Little, D.B. Rubin, Statistical Analysis with Missing Data, John Wiley & Sons, New York, 1987, Second edition published in 2002.
[3] R.O. Duda, P.E. Hart, D.G. Stork, Pattern Classification, 2nd edition, Wiley-Interscience, Hoboken, New Jersey, USA, 2001.
[4] C.M. Bishop, Pattern Recognition and Machine Learning, Springer, London, UK, 2006.
[5] Z. Ghahramani, M.I. Jordan, Supervised learning from incomplete data via an EM approach, in: J.D. Cowan, et al., (Eds.), Advances in Neural Information Processing Systems, vol. 6, Morgan Kaufmann Publishers Inc, San Mateo, CA, USA, 1994, pp. 120–127.
[6] P.K. Sharpe, R.J. Solly, Dealing with missing values in neural network-based diagnostic systems, Neural Comput. Appl. 3 (2) (1995) 73–77.
[7] J.R. Quinlan, Induction of decision trees, Mach. Learn. 1 (1) (1986) 81–106.
[8] R.J. Hathaway, J.C. Bezdek, Fuzzy C-means clustering of incomplete data, IEEE Trans. Syst. Man Cybern. B Cybern. 31 (5) (2001) 735–744.
[9] K. Pelckmans, J.D. Brabanter, J.A.K. Suykens, B.D. Moor, Handling missing values in support vector machine classifiers, Neural Netw. 18 (5–6) (2005) 684–692.
[10] A. Farhangfar, L. Kurgan, J. Dy, Impact of imputation of missing values on classification error for discrete data, Pattern Recognit. 41 (2008) 3692–3705.
[11] D.J. Mundfrom, A. Whitcomb, Imputing missing values: the effect on the accuracy of classification, in: Multiple Linear Regression Viewpoints, vol. 25, no. 1, 1998, pp. 13–19.
[12] G. Batista, M.C. Monard, A study of K-nearest neighbour as an imputation method, in: Proceedings of the Second International Conference on Hybrid Intelligent Systems, vol. 7, IOS Press, Amsterdam , Netherlands, 2002, pp. 251–260.
[13] J. Luengo, J.A. Saez, F. Herrera, Missing data imputation for fuzzy rule-based classification systems, Soft Comput. 16 (5) (2012) 863–881.
[14] D. Li, J.Deogun, W. Spaulding, B. Shuart, Towards missing data imputation: a study of fuzzy k-means clustering method, in: The Fourth International Conference of Rough Sets and Current Trends in Computing (RSCTC04), 2004, pp. 573–579.
[15] F. Fessant, S. Midenet, Self-organizing map for data imputation and correction in surveys, Neural Comput. Appl. 10 (4) (2002) 300–310.
[16] G. Shafer, A Mathematical Theory of Evidence, Princeton University Press, Press, 1976.
[17] F. Smarandache, J. Dezert (Eds.), Advances and Applications of DSmT for Information Fusion, vols. 1–4, American Research Press, Rehoboth, 2004–2015. ⟨http://fs.gallup.unm.edu/DSmT.htm.
[18] P. Smets, The combination of evidence in the transferable belief model, IEEE Trans. Pattern Anal. Mach. Intell. 12 (5) (1990) 447–458.
[19] A.-L. Jousselme, C. Liu, D. Grenier, E. Bossé, Measuring ambiguity in the evidence theory, IEEE Trans. Syst. Man Cybern. Part A 36 (September (5)) (2006) 890–903.
[20] H. Laanaya, A. Martin, D. Aboutajdine, A. Khenchaf, Support vector regression of membership functions and belief functions—application for pattern recognition, Inf. Fusion 11 (4) (2010) 338–350.
[21] T. Denœux, A k-nearest neighbor classification rule based on Dempster–Shafer Theory, IEEE Trans. Syst. Man Cybern. 25 (5) (1995) 804–813.
[22] 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.
[23] Z.-g. Liu, Q. Pan, J. Dezert, A new belief-based K-nearest neighbor classification method, Pattern Recognit. 46 (3) (2013) 834–844.
[24] Z.-g. Liu, Q. Pan, J. Dezert, G. Mercier, Credal classification rule for uncertain data based on belief functions, Pattern Recognit. 47 (7) (2014) 2532–2541.
[25] Z.-g. Liu, Q. Pan, G. Mercier, J. Dezert, A new incomplete pattern classification method based on evidential reasoning, IEEE Trans. Cybern. 45 (4) (2015) 635–646.
[26] 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.
[27] T. Denœux, A neural network classifier based on Dempster-Shafer theory, IEEE Trans. Syst. Man Cybern. A 30 (2) (2000) 131–150.
[28] X. Deng, Y. Hu, Felix T.S. Chan, S. Mahadevan, Y. Deng, Parameter estimation based on interval-valued belief structures, Eur. J. Oper. Res. 241 (2) (2015) 579–582.
[29] M.-H. Masson, T. Denœux, ECM: an evidential version of the fuzzy c-means algorithm, Pattern Recognit. 41 (4) (2008) 1384–1397.
[30] 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.
[31] Z.-g. Liu, Q. Pan, J. Dezert, G. Mercier, Credal c-means clustering method based on belief functions, Knowl.-Based Syst. 74 (2015) 119–132.
[32] K. Zhou, A. Martin, Q. Pan, Z.-g. Liu, Median evidential c-means algorithm and its application to community detection, Knowl.-Based Syst. 74 (2015) 69–88.
[33] T. Denœux, Maximum likelihood estimation from uncertain data in the belief function framework, IEEE Trans. Knowl. Data Eng. 25 (1) (2013) 119–130.
[34] 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.
[35] S. Huang, X. Su, Y. Hu, S. Mahadevan, Y. Deng, A new decision-making method by incomplete preferences based on evidence distance, Knowl.-Based Syst. 56 (2014) 264–272.
[36] X. Li, J. Dezert, F. Smarandache, X. Huang, Evidence supporting measure of similarity for reducing the complexity in information fusion, Inf. Sci. 181 (10) (2011) 1818–1835.
[37] D.q. Han, Y. Deng, C.z. Han, Sequential weighted combination for unreliable evidence based on evidence variance, Decis. Support Syst. 56 (2013) 387–393.
[38] T. Kohonen, The self-organizing map, Proc. IEEE 78 (9) (1990) 1464–1480.
[39] D. Dubois, H. Prade, Representation and combination of uncertainty with belief functions and possibility measures, Comput. Intell. 4 (4) (1988) 244–264.
[40] L.A. Zadeh, On the Validity of Dempster's Rule of Combination, Memo M79/24, 1979, University of California, Berkeley, USA.
[41] J. Dezert, A. Tchamova, On the validity of Dempster's fusion rule and its interpretation as a generalization of Bayesian fusion rule, Int. J. Intell. Syst. 29 (March (3)) (2014) 223–252.
[42] F. Smarandache, J. Dezert, Information fusion based on new proportional conflict redistribution rules, in: Proceedings of Fusion 2005, International Conference on Information Fusion, Philadelphia, PA, USA, July 25–29, 2005.
[43] I. Hammami, J. Dezert, G. Mercier, A. Hamouda, On the estimation of mass functions using Self Organizing Maps, in: Proceedings of Belief 2014 Conference, Oxford, UK, September 26–29, 2014.
[44] S. Geisser, Predictive inference: an introduction, Chapman and Hall, New York, NY, 1993.
[45] M. Lichman, UCI Machine Learning Repository, University of California, School of Information and Computer Science, Irvine, CA ⟨http://archive.ics.uci.edu/ml⟩, 2013.

Zhunga Liu was born in China, in 1984. He received the Bachelor and Master degree from Northwestern Polytechnical University (NPU), China, in 2007 and 2010. His Ph.D. degree is received from both Telecom Bretagne, France and NPU, in 2014. He has been an Associate Professor in NPU since 2014. His research work focuses on belief function theory and its application in data classification.

Quan Pan was born in China, in 1961. He received the Bachelor degree in Huazhong University of Science and Technology, and he received the Master and Doctor degrees in Northwestern Polytechnical University (NPU), in 1991 and 1997 respectively. He has been a Professor since 1998 in NPU. His current research interests are data classification and information fusion.

Jean Dezert was born in France, in 1962. He received the Electrical Engineering Degree from Ecole Française de Radioélectricité Electronique and Informatique (EFREI), Paris, in 1985, the D.E.A. degree in 1986 from the University Paris VII and his Ph.D. from the University Paris XI, Orsay, in 1990. Since 1993, he is Senior Research Scientist in the Information Modeling and Processing Department (DTIM) at ONERA. His current research interest focuses on belief functions theory and its applications, especially for DSmT which has been developed by him and Prof. Smarandache .

Arnaud Martin was born in France, in 1974. He received the Master and Ph.D. degrees from the University of Rennes I, Rennes, France respectively in 1998 and 2001. and his Habilitation à Diriger des Recherches from University of Occidental Britany, in 2009. Since 2010, he has been a Full Professor at the University of Rennes 1. His research interests mainly focus on belief function theory and pattern recognition.