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].