

Combination of supervised and unsupervised classification using the theory of belief functions

Fatma Karem, Mounir Dhibi, Arnaud Martin

Abstract In this paper, we propose to fuse both clustering and supervised classification approach in order to outperform the results of a classification algorithm. Indeed the results of the learning in supervised classification depend on the method and on the parameters chosen. Moreover the learning process is particularly difficult which few learning data and/or imprecise learning data. Hence, we define a classification approach using the theory of belief functions to fuse the results of one clustering and one supervised classification. This new approach applied on real databases allows good and promising results.

1 Introduction

Behind the term of classification, one distinguishes two types of classification: the supervised and unsupervised one. The unsupervised classification is also called *clustering*. In clustering, from given data representing some object, we try to find groups or *clusters* which are the most compact and separated as possible. Then, we can try to affect one of the found cluster to a new observed object [2]. Generally, we make such decision based on the analysis of the dispersion of the objects in the data set. In the supervised context, the process can also be divided in two steps: the learning one and the classification. The learning step build a discriminate function based on labeled data, an unknown

F. Karem, M. Dhibi
Research Unit PMI 09/UR/13-0, Zarouk College Gafsa 2112, Tunisie
e-mail: fatoumacy@yahoo.fr,mounir.dhibi@ensta-bretagne.fr

A. Martin
University of Rennes 1, UMR 6074 IRISA, rue Edouard Branly, BP 30219, 22302 Lannion
Cedex, France
e-mail: Arnaud.Martin@univ-rennes1.fr

information in clustering. From this function, in the classification step, a new observed object is affected to one of the classes given by the fixed labels. Whatever the type of classification, we face up to many problems. We are always looking for the appropriate method for a given problem without to be sure to achieve it. Indeed, the obtained results depend on the method and on parameters; the no-free lunch theorem assures us that there is no better algorithm. Therefore, the choice of the appropriate method and parameters is not easy for a given application. Furthermore in the supervised context, the learning data do not generally represent perfectly the real data we have to classify. For example, all real classes are not systematically well represented in the learning database. As a result, a possible solution to some of these classification problems is the fusion of clustering and supervised classification. The goal of this fusion is to reduce the imprecision of results by trying to make a compromise between both classifications.

Studying classification fusion approaches, most of them are dealing with the fusion of either supervised [9, 12] or unsupervised classification [3, 4, 11, 7, 8]. The unsupervised classification fusion approaches are more complex due to the absence of class labels: an association between the clusters coming from the different algorithms must be found. The researches made on the fusion between the clustering and the classification were used essentially in order to deploy the unsupervised in the learning of the supervised classification [6, 10, 13].

In this article, we propose a fusion approach combining supervised and unsupervised classification results. As framework, we choose the theory of belief functions which have been used with success to fuse supervised classification results [12]. This framework allows to represent the uncertainty and imprecision of the results of the clustering and supervised classification and to combine the results managing the conflict.

This paper is organized as follow: in the next section, we present the clustering and the supervised classification principles. In the third section, we explain the fusion based on the theory of belief functions. In section four, we present the proposed fusion approach and finally the last section presents the results given by an experimental study on real data.

2 Classification

The goal of the classification task is to identify the classes to which belong the objects representing by their characteristics or attributes. We distinguish two types of classification: supervised and unsupervised one.

Unsupervised classification or clustering

In the clustering, we want to group the similar objects of a population in clusters. Let's assume, we dispose of an ensemble of objects noted by $X = \{x_1, x_2, \dots, x_N\}$ characterized by an ensemble of descriptors D . Therefore, the data are D -multidimensional. The aim is to find the groups (or cluster) to which each object x belongs. Hereafter, the clusters are noted by $C = \{C_1, C_2, \dots, C_n\}$. The clustering can be formalized by a function noted by $Y_{\bar{s}}$, that associates each element of X to one or more elements of C . Generally, the clustering is essentially based on the dispersion analysis to find the real clusters. Many difficulties can arise in this task. The main difficulty is to find the borders of the clusters. To evaluate the results, we have to find some evaluation criteria measuring the quality of results. Usually, we use indexes called validity indexes. There is no standard or general index. Among the clustering methods, we mention for example K -means and the hierarchical classification.

Supervised classification

In the supervised context, the classification is based on two steps: the learning step and the classification step. In the learning step, we consider the objects in X already labeled, *i.e.* each object is associated to a known label belonging to an ensemble of classes noted by $\Theta = \{\theta_1, \theta_2, \dots, \theta_n\}$. This is the conceptual difference with the clustering. The goal of the learning step is to find the best discriminate function Cl associating each data of the learning database x using the descriptor set (noted by D) to the correct class in Θ . The classification step consists to predict the class of a new object based on the learning function. Among the classification methods, we mention the k -nearest neighbors (k -NN), the decision tree, the neural network, the support vector machine (SVM) [2]. In the supervised context, the lack of learning data or the availability of inappropriate one make problems. In this case, we can consider that the learning function to discriminate data is imprecise and uncertain and leads to bad results. The confusion matrices are generally used to evaluate supervised classification results.

In this paper, a new approach is proposed to overcome the classification problems identified previously. This approach is based on the fusion between the supervised classification and the clustering results using the theory of belief functions.

3 Information fusion using the theory of belief functions

The fusion of classifiers can be made in three levels of the classification process: data, characteristic and decision. The third level is the level of the classification results and is the most interesting for our study. Many frameworks have been used for information fusion, such as vote theory, theory of possibilities or theory of belief functions. The last one, also called Dempster-Shafer theory, allows to represent both imprecision and uncertainty through two functions: plausibility and belief. Both functions are derived from a function called mass function defined on all the subsets of the frame of discernment Θ , noted 2^Θ . That is the difference with the theory of probabilities where only singletons are considered. Let's design by m_j the mass function associated to the source S_j . The mass functions are defined on 2^Θ and affect a value from $[0, 1]$. Moreover, the mass functions verify the constraint:

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

Hence, the power set 2^Θ is the set of all the disjunctions of decisions θ_i in the classification context: $2^\Theta = \{\emptyset, \{\theta_1\}, \{\theta_2\}, \{\theta_1 \cup \theta_2\}, \dots, \Theta\}$. The decisions or classes θ_i must be exclusive but not necessarily exhaustive. The definition of mass functions depend on the context, but generic approaches can be used. We will use here a model based on probabilities proposed in [1]. There are many rules of combination in the theory of belief functions such as the conjunctive and the disjunctive one. The conjunctive combination, introduced by Dempster in its normalized form, combines the mass functions considering the intersections between the elements of 2^Θ [9, 12]. This combination is formulated as follows for M mass functions, $\forall A \in 2^\Theta$:

$$m(A) = \sum_{B_1 \cap B_2 \cap \dots \cap B_M = A} \prod_{j=1}^M m_j(B_j)$$

The obtained mass is the combination of the mass functions of each different sources. From this mass function, the decision to find the best class θ_i for the considered observation can be made with the pignistic probability. The pignistic probability is defined by:

$$bet(\theta_i) = \sum_{A \in 2^\Theta, \theta_i \in A} \frac{m(A)}{|A|(1 - m(\emptyset))} \quad (2)$$

where $|A|$ is the cardinality of A . This criterion is employed in a probabilistic context of decision. In the next section, we present the proposed approach to fuse the results of clustering and supervised classification.

4 Fusion of supervised classification and clustering results using the theory of belief functions

The most researches made in fusion are dealing with unsupervised classification (such as in [3]) or with supervised classification (such as in [12]). For the fusion of supervised classification and clustering was essentially done to deploy the unsupervised classification to make the learning of the supervised one [6, 10]. That is not the goal in this paper.

The proposed approach in this article fuses both types of classification to improve the results. Our approach is based on two main steps: the first one is to apply the clustering and the supervised classification on the learning database separately; the second step consists to fuse the results of classification approaches. Based on the two different outputs, we try to make a compromise between both classifiers. We must take into account the bad representation of the cluster's borders in clustering and the bad learning in supervised classification. We model that through the theory of belief functions. Therefore, as inputs of our process, we must define the mass functions of both sources: supervised and unsupervised classification. How to model these mass functions? First, to define the mass function for the supervised source, we choose the probabilistic model of Appriou [1] previously used with success.

Therefore, we define a mass function for each object x belonging to a class θ_j , we have n classes. We have for each class θ_j :

$$m_s^j(\theta_j) = \frac{\alpha_{sj} R_s p(\theta_j^f | \theta_i)}{1 + R_s p(\theta_j^f | \theta_i)} \quad (3)$$

$$m_s^j(\theta_j^c) = \frac{\alpha_{sj}}{1 + R_s p(\theta_j^f | \theta_i)} \quad (4)$$

$$m_s^j(\Theta) = 1 - \alpha_{sj} \quad (5)$$

We note by θ_j^f the class affected by the supervised classifier to the object x , by θ_i the real class and by α_{sj} the reliability coefficient of the supervised classification for the class θ_j^f . The conditional probabilities are estimated through the confusion matrix on the learning database:

$$\alpha_{sj} = \max p(\theta_j^f | \theta_i) \forall i = \{1, \dots, n\} \quad (6)$$

and

$$R_s = \max_{\theta_j^f} (p(\theta_j^f | \theta_i))^{-1} \quad (7)$$

For the unsupervised source, mass functions must also be defined on the discernment space Θ . However, the classes of Θ are unknown in clustering. We only dispose of clusters without any labels. Therefore the definition of

mass function is made by measuring the similarities between clusters and classes found by the supervised classification. If the found clusters are more similar to the classes, the clustering and supervised classification agree with each other. The similarity is calculated using recovery between clusters and classes. A class is considered similar to a cluster if it is recovered totally by the cluster. Therefore the biggest is the number of objects in common the biggest is the similarity. We look for the proportions of found classes $\theta_1^f, \dots, \theta_n^f$ by the supervised classifier in each cluster [4, 3]. $\forall x \in C_i$ with c the number of clusters found. The mass function for an object x to be in the class θ_j is as follows:

$$m_{ns}(\theta_j) = \frac{|C_i \cap \theta_j^f|}{|C_i|} \quad (8)$$

where $|C_i|$ is the number of elements in the cluster C_i and $|C_i \cap \theta_j^f|$, the number of elements in the intersection between C_i and θ_j^f . Then we discount the mass functions as follows, $\forall A \in 2^\Theta$ by:

$$m_{ns}^{\alpha_i}(A) = \alpha_i m_{ns}(A) \quad (9)$$

$$m_{ns}^{\alpha_i}(\Theta) = 1 - \alpha_i(1 - m_{ns}(\Theta)) \quad (10)$$

The discounting coefficient α_i depends on objects. We can not discount in the same way all the objects. An object situated in the center of cluster is considered more representative of the cluster than another one situated on the border for example. The coefficient α_i is defined as (v_i is the center of cluster C_i):

$$\alpha_i = e^{-\|x-v_i\|^2} \quad (11)$$

After calculating the mass functions for the two sources, we can combine using the conjunctive rule and we adopt as decision criterion the maximum of pignistic probability. Based on the construction of our mass functions for the non-supervised classifier, both mass functions cannot be considered cognitively independent. Other combination rules could be used. In our problem we look for known singletons thanks to the use of supervised classification. Each object is affected to a precise class. The pignistic probability is employed because we are in probabilistic context.

5 Experimental study

In this section we present the obtained results for our fusion approach between supervised classification and unsupervised classification. We conduct our experimental study on different databases coming from generic databases without missing values obtained from the U.C.I repository of Machine Learn-

ing databases. The aim is to demonstrate the performance of the proposed method and the influence of the fusion on the classification results. The experience is based on three unsupervised methods such as the fuzzy C-Means (FCM), the k -Means and the Mixture Model. For the supervised classification, we use the k -Nearest Neighbors and the Bayes Classifier. We show in the Tables 1 and 2 the obtained classification rates on the data before and after the fusion respectively for the k -NN with the FCM, the k -Means and the mixture model and the Bayes classifier with the FCM, the k -Means and the mixture model.

The number of clusters may be equal to the number given by the supervised classification or fixed by the user. The values shown in both tables 1 and 2 are obtained after cross-validation with ten trials of experiments. In each trial, we test with a test database taken from 10 databases. The fusion effect is remarkable in the table 1. In fact, we obtain a rate greater than 90% for the databases: iris, sensor-readings24 and haberman, a rate equal to 80% for breast-cancer and a rate about 60% for abalone database. In the table 2, we obtain a rate of 100% for iris, breast-cancer, sensor-readings24 and a rate greater than 70% for abalone and haberman. The error rate does not exceed 30% after fusion. We note that the obtained rate after fusion are better than before fusion.

Data	NbC	NbCl	NbA	CR-BF	CR-AF		
					FCM	k -Means	Mixture Model
Iris	3	3	5	96.67	100.00	100.00	100.00
Breast- Cancer wisconsin	2	2	11	64.52	80.00	80.00	80.00
Sensor-readings-24	4	4	5	84.00	100.00	100.00	100.00
Haberman	2	2	4	75.17	100.00	100.00	99.34
Abalone	2	2	8	53.10	61.70	61.27	59.42

Table 1 Results obtained with k -NN and FCM, k -Means and Mixture Model. NbC: number of classes, NbCl: number of clusters, NbA: Number of attributes, CR-BF: classification rate obtained before fusion, CR-AF classification rate obtained after fusion

Data	NbC	NbCl	NbA	CR-BF	CR-AF		
					FCM	k -Means	Mixture Model
Iris	3	3	5	95.33	100.00	100.00	100.00
Breast- Cancer wisconsin	2	2	11	96.00	100.00	100.00	100.00
Sensor-readings-24	4	4	5	52.57	100.00	100.00	100.00
Haberman	2	2	4	73.83	77.74	77.74	77.41
Abalone	2	2	8	51.95	73.08	73.59	66.62

Table 2 Results obtained with Bayes Classifier and FCM, k -Means and Mixture Model. NbC: number of classes, NbCl: number of clusters, NbA: Number of attributes, CR-BF: Classification rate obtained before fusion, CR-AF Classification rate obtained after fusion

6 Conclusion

This paper proposes a new approach allowing the fusion between supervised classification and clustering. Both methods have limits and problems. The fusion is established to improve the performance of the classification. We make the fusion with the belief function theory. The proposed approach showed encouraging results on classical and real databases. This work can be spread by studying results on imprecise and uncertain databases and on database with missing data. The final goal of this work is to apply the approach in very difficult applications such as sonar and medical images where the learning is difficult due to an incomplete knowledge of the reality.

References

1. A. Appriou, *Décision et Reconnaissance des formes en signal, Chapter "Discrimination multisérial par la théorie de l'évidence"*, Hermes Science Publication, 2002.
2. M. Campedel, *Classification supervisée*, Telecom Paris 2005.
3. G. Forestie, C. Wemmert et P. Gançarski, *Multisource Images Analysis Using Collaborative Clustering*, EURASIP Journal on Advances in Signal Processing, vol. 11, pp. 374-384, 2008.
4. P. Gançarski et C. Wemmert, *Collaborative multi-strategy classification: application to per-pixel analysis of images*, In Proceedings of the 6th international workshop on Multimedia data mining: mining integrated media and complex data, Vol. 6, pp. 595-608, 2005.
5. T. Denoeux, *A k-nearest neighbor classification rule based on Dempster-Shafer Theory*, IEEE Transactions on Systems, Man and Cybernetics, Vol.25, N. 5, pp. 904-913, 1995.
6. M. Guijarro et G. Pajares, *On combining classifiers through a fuzzy multi-criteria decision making approach: Applied to natural textured images*, Expert Systems with Applications, Vol. 39, pp. 7262-7269, 2009.
7. M. Masson et T.Denoeux, *Clustering interval-valued proximity data using belief functions*, Pattern Recognition, Vol. 25, pp. 163-171, 2004.
8. M. Masson et T.Denoeux, *Ensemble clustering in the belief functions framework*, International Journal of Approximate Reasoning, vol. 52, num. 1, pp. 92-109, 2011.
9. L. Xu, A. Krzyzak, et C.Y. Suen, *Methods of combining multiple classifiers and their applications to handwriting recognition*, IEEE Transactions on Systems, Man, and Cybernetics, Vol. 22, N. 3, pp. 418-435, 1992.
10. M. K. Urszula et T. Switek, *Combined Unsupervised-Supervised Classification Method*, In Proceedings of the 13th International Conference on Knowledge Based and Intelligent Information and Engineering Systems: Part II, Vol. 13, pp. 861-868, 2009.
11. C. Wemmert et P. Ganarski, *A Multi-View Voting Method to Combine Unsupervised Classifications*, In Proceedings of the 2nd IASTED International Conference on Artificial Intelligence and Applications, Vol. 2, pp. 447-453, 2002.
12. A. Martin, *Comparative study of information fusion methods for sonar images classification*, In Proceeding of the 8th International Conference on Information Fusion, Vol. 2, pp. 657-666, 2005.
13. Y. Prudent et A.Ennaji, *Clustering incrémental pour un apprentissage distribué : vers un système volutif et robuste*, In Conférence CAP, 2004.