FUZZY CLUSTERING WITH AMBIGUITY FOR MULTI-CLASSIFIERS FUSION : CLUSTERING-CLASSIFICATION COOPERATION

Veyis Gunes
Laboratoire L3i
Université de La Rochelle
Avenue Marillac
17042 La Rochelle Cedex (FRANCE)
vgunes@univ-lr.fr

Michel Ménard
Laboratoire L3i
Université de La Rochelle
Avenue Marillac
17042 La Rochelle Cedex (FRANCE)
mmenard@univ-lr.fr

Pierre Loonis
Laboratoire L3i
Université de La Rochelle
Avenue Marillac
17042 La Rochelle Cedex (FRANCE)
ploonis@univ-lr.fr

Summary

The main aim of this paper is to demonstrate the performance of multi-classifiers fusion based on fuzzy clustering with ambiguity. The problem is seen from the multi-decision point of view (i.e. several classification modules). Each classification module is specialized on a particular region of the features space. These regions are obtained by fuzzy clustering and constitute the original data set by union.

Keywords: Fuzzy clustering, Ambiguity, Fusion, Pattern recognition.

1 INTRODUCTION

Using a pattern recognition approach, the p parameters observed build up the pattern vector. The recognition is linked to the pattern classes to be discriminated in the $p-$dimensional representation space. Real world problems are so complex that a single classifier managing whole classes and features is not well-suited. Indeed, the training data set is too large for a correct learning. For the combination of multiple classifiers, an algorithm usually take two basic approaches :

- classifier fusion algorithms. In this case, individual classifiers are applied in parallel and their outputs are combined to achieve a consensus; frameworks of these algorithms include the vote theory [10], unanimous consensus [10] [2], polling methods which utilize heuristic decision rules [4], Bayes theory [10], Dempster-Shafer theory [8] [10], possibilistic theory [1], and multistage classification [3]. These methods often provide techniques for subdividing the features space into a suitable number of sub-spaces allowing the cooperation of different classification techniques. Even if the global behavior is improved due to a kind of meta-decision frontiers design, the specification of each module on a parameter sub-space decreases each classifier recognition rate [5].
- adaptive classifier selection. Each classification module is specialized on a specific region of features space. The methods attempt to predict which single classifier is most likely to be correct for a given sample. Thus, only the output of the selected classifier is considered in the decision step. Several papers present methods to select a classifier that use estimates of each individual classifier's local accuracy in small regions of features space surrounding an unknown test sample [9]. However, only the output of the best classifier for this data is used to make the final decision. These algorithms are often based on optimization methods, as for example, genetic algorithms and neural networks [7].

It is possible to combine both approaches using methods partitionning the features space into a suitable number of regions : the goal is to allow the cooperation of several classification modules. Moreover, in order to avoid a complex conflicts managing, the fusion step must be realized with only a subset of classifiers (the most concerned by).

In this paper, a pattern recognition approach is presented. This one is based on supervised and unsupervised procedures and uses the Dempster-Shafer theory of evidence. These procedures are combined to form an adaptive classifier selection and fusion system. The system starts with a given set of training patterns labeled among the known classes. The figures 1 and 2 illustrate the method. An unsupervised mechanism generates a new partition, including ambiguity clusters. This partition is obtained using the Fc+2M algorithm [6]. A classifier is associated with each simple cluster. The training set for each classifier is then defined on the cluster and its associated ambiguous clusters (see figure 2). Each classifier is trained separately. The decision making for an unknown pattern is the fusion of outputs from *neighbour* classifiers (see section 2.2). The method can be defined as follows :

- let $\Phi = (\phi_i)_{i\in[1,l]}$ be a family of l classes ; the recognition is linked to the pattern classes to be discriminated in the p-dimensional representation space ;
- let $X = (x_k)_{k\in[1,n]}$ be the family of n objects where $x_k = \left(x_{k1}, x_{k2}, \ldots, x_{kp}\right)^t$ is a pattern described by p features (i.e. $x_k \in R^p$). Let $X^{lr} \subset X$ be a set of training patterns labeled according to the known Φ family ;
- let $\Omega = (\omega_i)_{i\in[1,c]}$ be a family of c clusters; These clusters are given by the fuzzy clustering processing applied in the p-dimensional representation space. Since there are 2^c subsets in a set of c elements, we obtain $2^c - 1$ regions, excluding the empty set, in a $c-$cluster problem. We attempt to partition the features space into regions; each of these corresponds to a subset of X^{lr}. Thus, in a $c-$cluster problem there are c simple clusters, $(\{\omega_i\})_{i\in[1,c]}$, and to exercise the ambiguity reject option, we associate an ambiguity cluster with each subset of clusters $A \in 2^\Omega \setminus \{\emptyset, (\{\omega_i\})_{i\in[1,c]}\}$. In figure 3 there are $2^3 - 1$ regions corresponding to the subsets $\{\omega_1\}$, $\{\omega_2\}$, $\{\omega_3\}$, $\omega_{12} = \{\omega_1, \omega_2\}$, $\omega_{13} = \{\omega_1, \omega_3\}$, $\omega_{23} = \{\omega_2, \omega_3\}$ and $\omega_{123} = \{\omega_1, \omega_2, \omega_3\}$;
- let $E = (e_k)_{k\in[1,c]}$ be a set of c classifiers computing the possibility that an unknown test sample belongs to the classes $(\phi_i)_{i\in[1,l]}$. A classifier is associated with each cluster i.e. the training data set, S_{ω_i} for the classifier e_i is defined on the cluster ω_i and its associated ambiguous clusters (see figure 2).

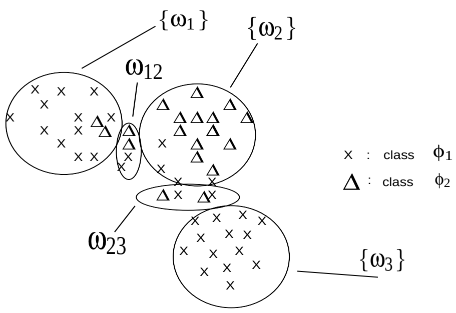

Figure 1: The partition of the sets of training patterns, $X^{lr} \subset X$, if the ambiguity reject is exercised. $\{\omega_1\}$, $\{\omega_2\}$, $\{\omega_3\}$: simple clusters; ω_{12}, ω_{23} : ambiguous clusters.

Each classification module is specialized on a specific region of the features space. These regions are obtained by fuzzy clustering and produce the original data set by union. The goal is to reduce the amount of data by merging similar patterns together into clusters. Moreover, we introduce the ambiguity notion in clustering, which deals with patterns lying near the clusters boundaries. The overall system is parallel, since different classifiers work with their own

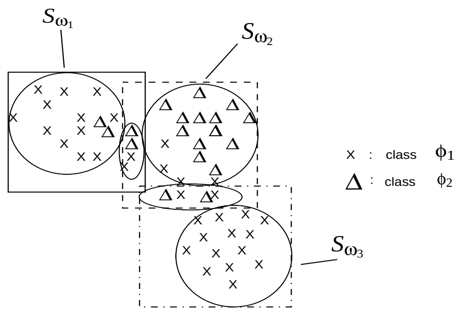

Figure 2: One classifier is associated with each cluster. The training data set, S_{ω_i} for the classifier e_i is defined on the cluster ω_i and its associated ambiguous clusters.

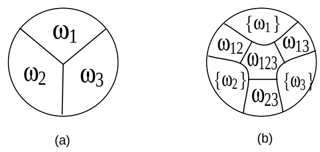

Figure 3: (a) the partition of the pattern space X into three regions corresponding to simple cluster. (b) the partition if the ambiguity reject is active. We associate an ambiguity cluster with each subset $A \in 2^\Omega \setminus \{\emptyset, (\{\omega_i\})_{i\in[1,3]}\}$.

training data clusters. The algorithm makes it possible the adaptive classifiers selection in the sense that the fuzzy clustering with ambiguity gives adapted training data regions of features space. The decision making is the fusion of outputs from *neighbour* classifiers.

2 A PARALLEL AND ADAPTIVE MULTI-CLASSIFIERS FUSION

The overall system consists of four major components : a set of c classifiers $(e_k)_{k\in[1,c]}$ which are responsible for calculating the possibility that a particular test pattern belongs to cluster ϕ_i, $i \in [1, l]$, a fusion module $\mathcal{F}$ which merges the outputs from subset of individual classifiers, a decision module $\mathcal{D}$ which assigns the test pattern to one of the l classes and a selection module which chooses the subset of classifiers from the set of classifiers for a given pattern.

2.1 THE CLUSTERING PROCEDURE : SELECTION MODULE

The Fc+2M algorithm generates a new partition $\mathcal{P}$ of X^{lr}, including ambiguity clusters. The number of simple clusters, c, is generally higher than the number of classes, l. The use of clustering on complex and noised data permits to specify the training sets and the neighbourhood of the classifiers, and deals with the cooperation between the classifiers allowing the multiple points of view decision management. The main idea is that the clustering procedure can manage the selection of the most adapted classifiers.

2.2 THE CLASSIFIERS

The bayesian classifier, well known statistical approach, is considered in this study. The main hypothesis is the knowledge on the distribution law whose features are determined on the learning set, X^{lr}. One classifier is associated with each simple cluster. The training data set, S_{ω_i}, for the classifier e_i is defined on the cluster ω_i and its associated ambiguous clusters : $S_{\omega_i} = \{x_k \in A \ : \forall A \in 2^{\Omega}$ and $\omega_i \in A\}$. The decision classes for e_i are defined by $\{\phi_j \in \mathbf{\Phi} \ : \phi_j \ni x_k \, \forall x_k \in S_{\omega_i}\}$. Let $(e_k)_{k \in [1,c]}$ be the set of classifiers. The set of neighbouring classifiers for the classifier e_i is defined by: $E_i = \{e_k \ : S_{\omega_i} \cap S_{\omega_k} \neq \emptyset\}$.

2.3 FUSION MODULE

Information fusion is an important aspect of any intelligent system. Let $X^t \subset X (X^t \cap X^{lr} = \emptyset)$ be the set of unknown patterns. For classification, an unknown pattern $x_k \in X^t$ is assigned to the partition $\mathcal{P}$. Thus the following steps for pattern recognition at this level are :

- if x_k is classified in a simple cluster, ω_i, the output of the associated classifier, e_i, with this cluster is used to make the final decision;
- if x_k is classified in an ambiguous cluster, the decision making is expressed as an aggregation of the set of outputs from neighbour classifiers (see figure 4) such as : $\{e_j \ : \omega_j \in A\}$

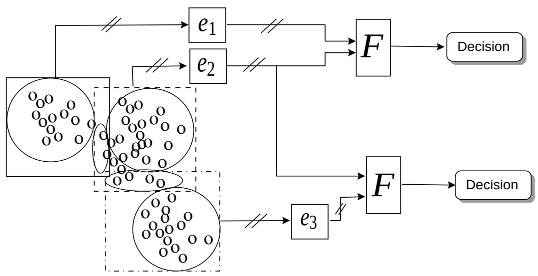

Figure 4: The decision making for an unknown pattern is expressed as an aggregation of the set of outputs from neighbour classifiers.

Several methods for combining pieces of information issued from several sources, based on certainty factors as MYCIN algorithm, or on credibility and plausibility measurements as in Evidence Theory were proposed in literature. Our proposed Fusion module is based on the combination using the Dempster-Shafer's orthogonal rule [8].

3 RESULTS

The above method, named 3C (as clustering-classification cooperation), is tested on a set of synthetic data (2-dimensional data). The learning set X is composed of 210 patterns (figure 5). Our choice is $\sigma = 2$ (standard deviation).

- Φ_1 : 100 Gaussian distributed patterns according to :

$$m_1 = \begin{pmatrix} 10 & 20 \end{pmatrix}^t, \Sigma_1 = \begin{pmatrix} \sigma^2 & 0 \\ 0 & \sigma^2 \end{pmatrix}$$

- Φ_2 : $(\Sigma_2 = \Sigma_1)$
  - Gaussian distributed, according to :

$$m_{2a} = \begin{pmatrix} 10 & 10 \end{pmatrix}^t \, 50 \, patterns$$

$$m_{2b} = \begin{pmatrix} 10 & 30 \end{pmatrix}^t \, 50 \, patterns$$

  - 10 patterns (on the right of Φ_1)

The test set is generated in the same conditions, as shown in figure 6. The Euclidian metric is used as a distance measure. Let's notice that during the clustering stage (Fc+2M), distance rejection is avoided. The results obtained are shown in figure 7 (c=2). Fc+2M parameters are : $m = 1.5$ and $\alpha = 0.1$. The partition obtained is as shown in table 1 for $c = 2$ and in table 2 for $c = 3$. All the values in the tables are expressed in percentage. The results after fusions are shown in table 3 (for $c = 2$). A comparison, with the bayesian theory of decision applied to the whole of data set, is performed. Table 4 shows the results for the three methods (bayesian, 3C algorithm with $c = 2$ and $c = 3$). The overall best performance is obtained with the 3C method for c=2.

Table 1: Fc+2M partition of X^t for $c = 2$: Φ_1 and Φ_2 are real classes; ω_1 and ω_2 are simple clusters; ω_{12} is the ambiguous cluster.

$\nearrow$	ω_1	ω_2	ω_{12}
Φ_1	48.0	15.0	37.0
Φ_2	50.91	49.09	0

Table 2: Fc+2M partition of X^t for $c = 3$: There are 3 simple clusters and 4 possible ambiguous clusters (2 of them, ω_{12} and ω_{13} exist really).

$\nearrow$	ω_1	ω_2	ω_3	ω_{12}	ω_{13}	ω_{23}	ω_{123}
Φ_1	67.0	0	0	17.0	16.0	0	0
Φ_2	3.6	36.4	45.5	11.8	2.7	0	0

Table 3: 3C Algorithm : Confusion matrix for $c = 2$.

$\nearrow$	Φ_1	Φ_2
Φ_1	95.0	5.0
Φ_2	0.91	99.09

Table 4: Comparison with bayesian classifier.

	Bayes	3C (c=2)	3C (c=3)
Success	88.6	97.1	96.7
Confusion	11.4	2.9	3.3

4 CONCLUSION

In this paper, we have shown that the ambiguity reject in the clustering stage is very useful in order to constitute subsets of data for the learning stage of more localized set of bayesian classifiers. Our results show that in the case of non-gaussian classes, a much better classifier than a single global bayesian classifier can be obtained by using multiple local bayesian classifiers associated with fusion by Dempster-Shafer orthogonal rule. In this latter stage, ignorance could be calculated by using rates of patterns affected to each single classifier.

References

[1] D. Dubois and H. Prade. Possibility theory. *Plenum Press. New York*, 1988.

[2] T. K. Ho, J.J. Hull, and S.N. Srihari. Decision combination in multiple classifier systems. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 16:66–75, jan 1994.

[3] Y.S. Huang and Suen C.Y. A method of combining multiple classifiers - a neural network approach. *Proc. 12th Int'l Conf. Pattern Recognition and Computer Vision. Jérusalem*, pages 473–475, 1994.

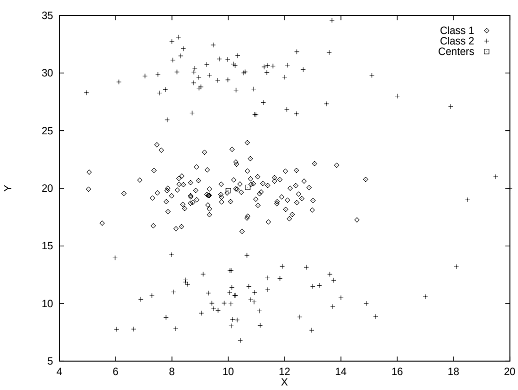

Figure 5: *Learning set:* $X^{lr} \subset X$

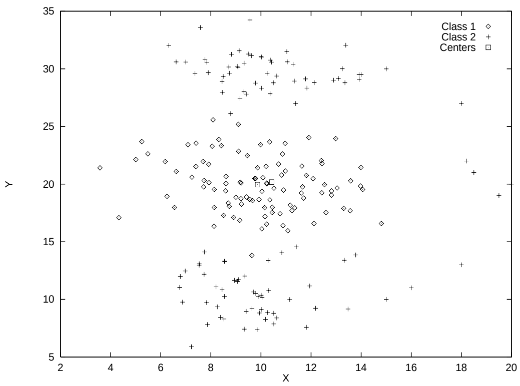

Figure 6: *Test set:* $X^t \subset X$

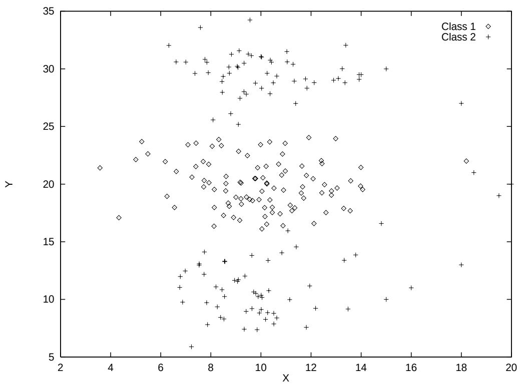

Figure 7: *Results obtained with 3C algorithm*

[4] F. Kimura and M. Shridar. Handwritten numerical recognition based on multiple algorithms. *Pattern recognition*, 24(10):969–983, 1991.

[5] P. Loonis and M. Ménard. z_{01}, a reliability limit in pattern recognition: Application in exterior scenes area and automatic fish sorting. In *Proceedings of IPMU*, Paris, july 1998.

[6] M.Menard, C.Demko, and P.Loonis. The fuzzy c+2-means: Solving the ambiguity reject in clustering. In *Pattern recognition. To be published*.

[7] P.Loonis, M.Ménard, and C.Demko. A new genetic algorithm for the multi-classifiers fusion optimization. In *Proceedings of IPMU*, Granada, Spain, July 1996.

[8] G. Shafer. A mathematical theory of evidence. *Princeton University Press,Princeton, N.J.*, 1976.

[9] K. Woods, W.P. Kegelmeyer, and JR.K. Bowyer. Combination of multiple classifiers using local accuracy estimates. In *PAMI*, April 1997.

[10] L. Xu, A. Krzyzak, and C.Y. Suen. Methods of combining multiple classifiers and their applications to handwritting recognition. *IEEE Trans. Systems Man and Cybernetics*, 22(3):418–435, 1992.