

A Pretopological Approach for Pattern Classification with Reject Options

Carl Frélicot¹ and Hubert Emptoz²

¹ Université de La Rochelle, Laboratoire d'Informatique et d'Imagerie Industrielle,
17042 La Rochelle Cedex 1, FRANCE

Phone: +33 546 458 234, Fax: +33 546 458 242, e-mail: cfrelico@gi.univ-lr.fr

² I.N.S.A. de Lyon, Laboratoire de Reconnaissance des Formes et Vision,
69621 Villeurbanne Cedex, FRANCE

Abstract. In this paper, we present a pretopological approach for pattern classification with reject options. The pretopological approach, based on growing ϵ -neighborhoods, already has proved its efficiency in reducing computation time and storage requirements compared to a k -Nearest Neighbors approach.

By including ambiguity and distance reject options, we give to such an approach more adaptability to real classification problems for which classes generally are not clearly separable and/or not completely known. In order to control both types of rejection, the proposed classifier needs a unique parameter to be fixed whereas two parameters generally are necessary (one for each reject type). We also have observed that the behavior of the classifier (depending of the parameter value) with respect to both kinds of rejection is similar to the behavior of well-known rejection-based classifiers proposed so far in the literature.

1 Introduction

Neighborhood-based classifiers, e.g. the k - *Nearest Neighbors*, are attractive for the pattern classification problem because they are local densities estimators. Their main drawbacks are their computation time and storage requirements. By using a condensed version, these limits can be reduced (see [7]). The pretopological approach for pattern classification proposed in [4] is based on the same idea. Furthermore, the concept of *adhesion* it is based on expresses a perceptive point of view which is quite natural.

But it is well known that classical classifiers often fail when facing real applications because they do not take into account neither the impreciseness of the classes reflected by overlapping areas nor their non completeness. Rejection-based classifiers aim at dealing with these two kinds of uncertainty, e.g. in [3], [5]. The classifier we present is a revisited version of the pretopological-based one in the framework of classification with reject options.

2 Classification with Reject Options

We address here the classification problem which can be defined as follows. Let $x = (x_1, x_2, \dots, x_p)^t$ be a pattern described by p features and $\{\omega_1, \omega_2, \dots, \omega_c\}$ a set of c classes. According to many authors, any function $D : \mathcal{R}^p \rightarrow L$ a label vectors

space, $x \mapsto l(x) = (l_1(x), l_2(x), \dots, l_c(x))^t$ is a classifier. Supervised classifier design consists in identifying the parameters of D from a learning set X , i.e. a set of patterns whose labels are known. In particular, D is said to be a *crisp* classifier if $L = \{l : l_i(x) \in \{0, 1\}, \sum_{i=1}^c l_i(x) = 1\}$, resulting in partitioning the feature space \mathbb{R}^p into c mutually exclusive areas $\{\Omega_1, \Omega_2, \dots, \Omega_c\} = \Omega$, as shown in Figure 1. So built up classifiers may not be very efficient from a practical point of view because:

1. areas $\Omega_i \subset \Omega$ are open whereas learning subsets $X_i \subset X$ are finite
2. areas boundaries are sharp whereas learning subsets may partially overlap

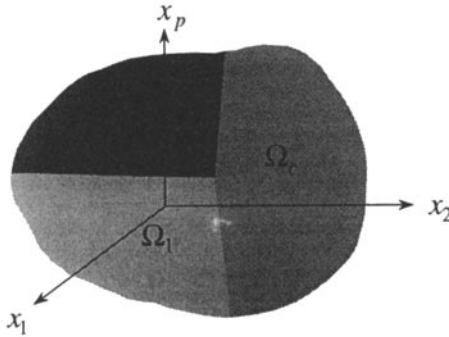


Fig. 1. Classes areas

In order to overcome these limits, we rather to design rejection-based classifiers. Such classifiers allow:

1. not to classify a pattern x in any known class if x lies far from all the patterns in X ; this concept is known as *distance rejection* and the classifier must allow to assign the label space origin to such a rejected pattern
2. to classify a pattern x in several classes if x lies where part of the learning subsets overlap; this concept is known as *ambiguity rejection* and the classifier must allow to assign non exclusive labels to such a rejected pattern

Therefore we can define a *rejection-based hard classifier* as any function $D : \mathbb{R}^p \rightarrow L = \{l : l_i(x) \in \{0, 1\}, 0 \leq \sum_{i=1}^c l_i(x) \leq c\}$. Such classifiers result in partitioning the feature space \mathbb{R}^p into $(c + 2)$ areas $\{\Omega \cup \Omega_0 \cup \Omega_a\}$, as shown in Figure 2, where:

1. $\Omega = \{\Omega_1, \Omega_2, \dots, \Omega_c\}$, with Ω_i being the closed area associated to class ω_i
2. $\Omega_0 = \bigcup_i \Omega_i$ the complementary set of classes areas in \mathbb{R}^p , associated to an additional class say ω_0 , i.e. the distance reject class
3. $\Omega_a = \bigcup_{i,j \neq i} (\Omega_i \cap \Omega_j)$, all the overlaps, associated to another additional class say ω_a , i.e. the ambiguity reject class

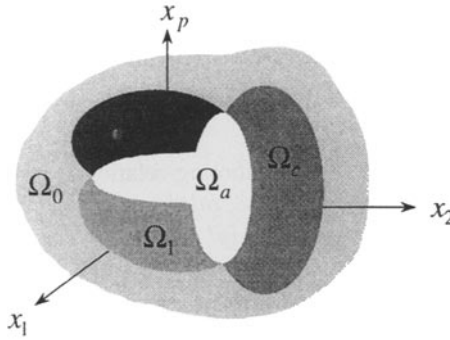


Fig. 2. Rejection-based classes areas

3 Pretopology-Based Classification

3.1 Pretopology and Neighborhood

From a perceptive point of view, a pattern $x \in X$ is rarely seen as an individual but there are neighbors it is not dissociated from. A part ω of X is rarely dissociated from its dilated part as well. In order to deal with this perceptive concept, the *adhesion* has been defined [4] as follows:

Definition 1. Let $P(X)$ be the power set of X . We shall call *adhesion* any function $ad : P(X) \rightarrow P(X)$ satisfying:

1. $ad(\emptyset) = \emptyset$
2. $\omega \subset ad(\omega) \quad \forall \omega \in P(X)$

Related definitions are:

Definition 2. (X, ad) is called a *pretopological space*.

Definition 3. $\omega \in X$ is called a *closed set* if and only if $ad(\omega) = \omega$.

It is possible to compose the adhesion application as follows: $x = ad^0(x)$, $ad(x) = ad^1(x)$, ..., $ad(ad^n(x)) = ad^{n+1}(x)$. In $ad^j(x)$, j will be called the *adhesion order*. Note *non idempotency* of ad allows to propagate adhesions ; this will be of major importance for the classes areas learning process.

A neighborhood is a binary relation defined on X which can generate pretopologies. Let $V(x)$ be the set of patterns x' neighboring x in X and $V^{-1}(x)$ the set of patterns x' including x in their neighborhood. Associated with X , each of the following adhesions forms a pretopological space:

1. $ad(\emptyset) = \emptyset$
 $ad(x) = x \cup V(x) \quad \forall x \in X$
 $ad(\omega) = \bigcup_{x \in \omega} ad(x) \quad \forall \omega \in X$

2. $ad^{-1}(\emptyset) = \emptyset$
 $ad^{-1}(x) = x \cup V^{-1}(x) \quad \forall x \in X$
 $ad^{-1}(\omega) = \bigcup_{x \in \omega} ad^{-1}(x) \quad \forall \omega \in X$
3. $ad \cap ad^{-1}$

It is worthy of note that if the neighboring relation is symmetrical then $V(x)$ and $V^{-1}(x)$ are identical and therefore ad , ad^{-1} and $ad \cap ad^{-1}$ generate the same pretopology. This is the case for the $\epsilon - Neighborhood$, we use:

Definition 4. Given any distance metric d in \mathfrak{R}^p and $\epsilon \in \mathfrak{R}$, the set $V_\epsilon(x) = \{x' \in X, d(x, x') \leq \epsilon\}$ is called a $\epsilon - Neighborhood$ of x .

Such a neighborhood generates the pretopology defined by the following adhesion:

$$ad(\emptyset) = \emptyset$$

$$ad(x) = x \cup V_\epsilon(x) \quad \forall x \in X$$

$$ad(\omega) = \bigcup_{x \in \omega} ad(x)$$

3.2 Learning Process

We aim at designing a classifier D using a $\epsilon - N$ pretopology defined on a learning set $X = \bigcup_{i=1}^c X_i$, where $X_i = \{x \in X : l_i(x) = 1, l_{j \neq i}(x) = 0\}$, i.e. the set of patterns from X whose labels correspond to class ω_i . The learning process consists in :

- $\forall i = 1, c$
1. $\forall x \in X_i$, finding the highest order $n(x)$ such as:
 - (a) $ad^{n(x)}(x) \cap \overline{X_i} = \emptyset$
 - (b) $ad^{n(x)+1}(x) \cap \overline{X_i} \neq \emptyset$
 where $\overline{X_i}$ denotes the complementary set of X_i in the learning set X ,
 $ad^{n(x)}(x)$ is approximated by $V_\epsilon(x) = \{x' \in X_i, d(x, x') \leq \epsilon\}$ whatever the distance metric d is,
 and $\epsilon(x) = \max_{x' \in ad^{n(x)}(x)} d(x, x')$ is called the *radius* of $ad^{n(x)}(x)$.
2. defining $\Omega_i = \bigcup_{x \in X_i} ad^{n(x)}(x)$ as the area of the feature space associated with class ω_i .

Note $\bigcup_{x \in X_i} ad(x) \subset \bigcup_{x \in X_i} ad^{n(x)}(x)$. This means that adhesions of lower order than $n(x)$ may be explicitly eliminated. The class ω_i parameters will then consist of all the couples $(x, \epsilon(x))$ needed to define the class area Ω_i . Some of such couples may correspond to overlapping adhesions. By sorting the couples in the descending order of their radius $\epsilon(x)$, an adhesion which is completely included in the union of different adhesions of higher radius (from the same class) will be implicitly eliminated because it will never be taken into account by the classifier. Therefore, the final step of the learning process is:

3. sort all $ad^{n(x)}(x), x \in X_i$ by descending order of $\epsilon(x)$

A second learning strategy could consist in eliminating zero-order adhesions because they correspond to learning patterns whose neighborhood is composed of patterns issued from different class(es).

Let us illustrate these learning processes on the very simple learning set $X = X_1 \cup X_2$ of Figure 3 in \mathfrak{R}^2 . Without loss of generality, the identity metric has been chosen. Then the distance between two patterns from X is defined by $d^2(x, x') = (x - x')^t(x - x')$ resulting to hyperspherical adhesions. Figure 4 shows the $ad^{n(x)}(x)$ that are necessary for defining both classes areas Ω_1 and Ω_2 . In these figures, white symbols correspond to patterns whose adhesion order $n(x)$ equals zero ; the corresponding couples $(x, \epsilon(x))$ will be removed if the second learning strategy is chosen. Black ones correspond to patterns that are sufficient to define the classes areas, whereas grey symbols have been associated to patterns whose adhesion will never be taken into account by the classifier.

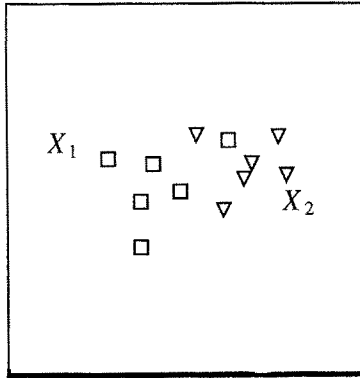


Fig. 3. Learning set $X_1 \cup X_2$

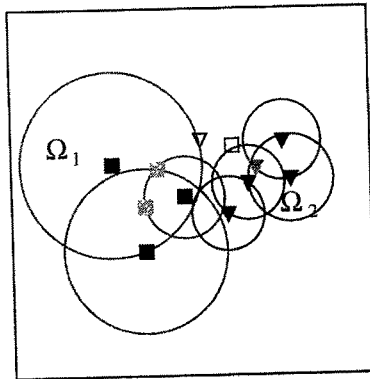


Fig. 4. Classes areas $\{\Omega_1, \Omega_2\}$

3.3 Rejection-Based Classifier

The classes areas learned as above do partition the feature space \mathfrak{R}^p into $(c + 2)$ areas as a rejection-based classifier does: $\{\Omega_1, \Omega_2, \dots, \Omega_c\} \cup \Omega_0 \cup \Omega_a$. The rejection-based classification algorithm we propose is the following one:

- $\forall x'$ to be classified
1. $\forall i = 1, c$
 seek for a $(x, \epsilon(x))$ in the sorted Ω_i -list such as $d(x, x') \leq \epsilon(x)$
 set $l_i(x') = 1$ if found, $l_i(x') = 0$ otherwise
 2. if $\sum_{i=1}^c l_i(x) \begin{cases} < 1 \\ = 1 \\ > 1 \end{cases}$ x' is classified into class $\begin{Bmatrix} \omega_0 \\ \omega_{\text{argmax}_{j,l_j}(x)} \\ \omega_a \end{Bmatrix}$

It is worthy of note that, as mentioned previously, all the patterns $x \in X$ are not considered (except for the distance reject case). This means a lower amount of computation time compared to other neighborhood-based classification methods, e.g. the k - *NearestNeighbors* for which all the distances $d(x', x)$ ($\forall x \in X$) have to be computed and sorted. On the other hand, the learning process These properties of pre-topological approaches are pointed out in [6].

A major advantage of the proposed classifier is that the distance reject class area Ω_0 is completely defined by the learned lists $(x \in X_i, \epsilon(x))$. We mean it do not require any external parameter for distance rejection as other classifier do, e.g. a mixture density threshold for the parametric approach in [3] or a distance threshold for the rejection-based k -*NN* classifier in [2]. The same remark applies to the ambiguity reject class area Ω_a . No external parameter controls ambiguity rejection as a reject cost for the so-called *Chow's rule* [1] or an additional number k' of neighbors for the the so-called (k, k') - *NN* rule in [9].

However, the counterpart is that such a scheme may lead either to misclassification or to distance rejection of patterns lying where different classes overlap. This is due to the definition of adhesions which have low radii in the overlapping areas. This problem can be reduced by allowing to define, for each pattern from the learning set, adhesions including a fixed rate of foreigner patterns, i.e. patterns issued from other classes, during the learning process. Only the first step has to be replaced by :

1. $\forall x \in X_i$, find the highest order $n(x)$ such as:
 - (a) $|ad^{n(x)}(x) \cap \overline{X_i}| \leq \alpha_i |ad^{n(x)}(x) \cap X_i|$
 - (b) $|ad^{n(x)+1}(x) \cap \overline{X_i}| \geq \alpha_i |ad^{n(x)}(x) \cap X_i|$
 where $||$ denotes the cardinality operator.

Note parameters $0 \leq \alpha_i \leq 1$, which control ambiguity rejection, may not be class-dependant, i.e. $\alpha_i = \alpha \forall i = 1, c$. The previous approach is obtained using $\alpha_i = 0 \forall i = 1, c$. This refinement obviously results in decreasing the number of adhesions to be taken into account by the classifier and therefore decreasing the hypervolume of Ω_0 and increasing Ω_a 's one. Therefore, only one parameter (α) is necessary to control both

kinds of rejection where two parameters are generally required. It is also worthy of note that the proposed classifier is *class-selective* ambiguity rejection-based as in [8] whereas the ones mentioned above are not, because a pattern x' associated with class ω_a is ambiguity rejected between the (sub)set of classes such as $l_i(x') = 1$; this might be of great importance in some applications.

4 Experimental Results

We do not intend to show the advantages of the proposed classifier compared to other neighborhood-based ones, this has been already done in [4] and [6]. We rather to focus on its ability to deal with both kinds of rejection, using the second learning strategy. In particular, we are interested in the role of parameter α in controlling the reject rates.

Figures 5 and 6 show the classification areas, including reject ones, obtained from the artificial learning set of Figure 3 with $\alpha = 0$ and $\alpha = 0.3$ respectively. We can see that since the classes areas Ω_i grow with α , the ambiguity reject one Ω_a grows as well whereas the distance reject one Ω_0 grows shorter. Due the second learning strategy, the two learning patterns whose adhesion order were zero lie in Ω_0 with $\alpha = 0$ and lie in Ω_a with $\alpha = 0.3$.

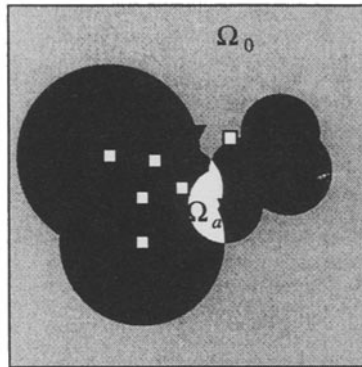


Fig. 5. Classification areas ($\alpha = 0$)

We also have tested the classifier on a real data set, the well-known *Iris* data set consisting of 150 patterns described by $p = 4$ features, divided in $c = 3$ classes of 50 patterns each. Two classes overlap whereas the third one is well separated. Because of learning computation time, we performed a resubstitution test, i.e. the same data set is used for both learning and classification processes. Figure 7 shows the obtained probabilities estimates: \hat{P}_c (correct classification), \hat{P}_a (ambiguity rejection), \hat{P}_e (error or misclassification) \hat{P}_0 (distance rejection) as functions of the parameter α which controls the reject rates. As expected, \hat{P}_a increased with α whereas \hat{P}_c , \hat{P}_e and \hat{P}_0 decreased, i.e. the more foreign patterns accepted in adhesions, the larger the adhesions are and

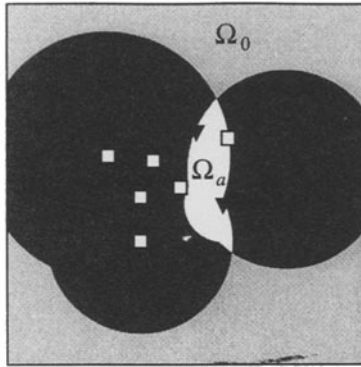


Fig. 6. Classification areas ($\alpha = 0.3$)

therefore the more ambiguity rejection, the less misclassification and distance rejection happen. It clearly means that α behaves: like a distance threshold as in [2] or inversely like a mixture threshold as in [3] with respect to \hat{P}_0 , and inversely like a reject cost as in [1] or like a number of neighbors k' as in [9] with respect to \hat{P}_a and \hat{P}_e .

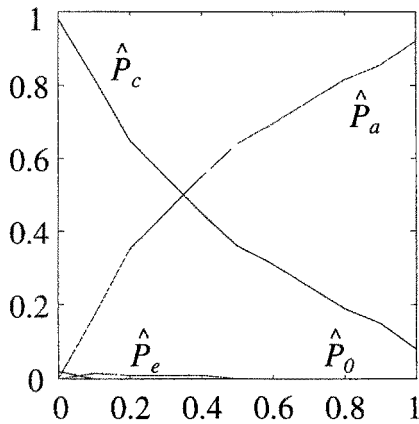


Fig. 7. Estimated probabilities vs α

5 Conclusion

In this paper, we have presented a classifier based on a ϵ -neighborhood pretopological approach in the framework of rejection-based classification. This pretopological approach allows to reduce the computation and storage limits of classical neighborhood-based methods. The proposed classifier becomes very attractive when reject principles are

included. Thus, ambiguous and/or non representative patterns can be rejected instead of being misclassified. The resulting classifier behave like most of rejection-based classification methods. Its main characteristic is that both kinds of rejection can be controlled using a unique parameter, whereas two parameters are generally needed.

References

1. Chow, C.K.: An optimum character recognition system using decision functions. *IEEE Transactions on Information Theory* **16** (1970) 41-46
2. Dasarathy, B.V.: Noising around the neighborhood: a new system structure and classification rule in partially exposed environment. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **2** (1980) 67-71
3. Dubuisson, B., Masson, M.H.: A statistical decision rule with incomplete knowledge about classes. *Pattern Recognition* **26** (1993) 155-165
4. Lebourgeois, F., Emptoz, H.: Pretopological approach for supervised learning. *Proceedings 13th International Conference on Pattern Recognition* (1996) 256-260
5. Frélicot, C.: Multiprototype-based fuzzy classification and reject options. *Proc. 5th IEEE International Conference on Fuzzy Systems* (1996) 2026-2031
6. Frélicot, C., Lebourgeois, F.: A pretopology-based supervised pattern classifier. *Proc. 14th International Conference on Pattern Recognition* (1998)
7. Fukunaga, K.: *Introduction to statistical pattern recognition*. Academic Press (1990) Second edition
8. Ha, T.M.: The optimum class-selective rejection rule. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **19** (1997) 608-615
9. Hellman, M.E.: The nearest neighbor classification rule with a reject option. *IEEE Transactions on Systems Man and Cybernetics* **6** (1970) 179-185