Discrimination basée sur un critère d'homogénéité locale

Discrimination by optimizing a local consistency criterion

Abdelkader ZIGHED
URA 934 Bt 101 Université Lyon I,
43 Bd du 11 Novembre 1918,
69100 VILLEURBANNE

Docteur Ingénieur (Université Lyon 1 1985), maître de conférences en informatique à l'université Lumière Lyon 2 et membre de l'Unité Associée 934 du CNRS et de l'Université C. Bernard Lyon 1. Il a réalisé de nombreux travaux de rcherche sur la reconnaissance des formes. Il est, notamment, l'auteur de SIPINA, une méthode de discrimination particulièrement adaptée au traitement des petits échantillons. Ses recherches actuelles s'orientent vers les techniques de géométrie informatique et leur utilisation dans les problèmes d'apprentissage supervisé.

Daniel TOURNISSOUX
URA 934 Bt 101 Université Lyon I,
43 Bd du 11 Novembre 1918,
69100 VILLEURBANNE

Docteur es Sciences (Université Lyon 1, 1980), Professeur à l'université Lyon 1 et membre de l'Unité Associée 934 du CNRS et de l'Université C. Bernard Lyon 1. Ses travaux de recherche ont porté sur la théorie de l'information et des problèmes d'aides au diagnostic. Il travaille sur les problèmes de restauration d'images et des méthodes de discrimination non paramétriques dans des espaces de représentation à structure continue.

Jean-Paul AURAY
URA 934 Bt 101 Université, Lyon I,
43 Bd du 11 Novembre 1918,
69100 VILLEURBANNE

Docteur es Sciences (Université Lyon 1, 1983), Directeur de recherche CRNS. A travaillé dans le développement des modèles mathématiques appliqués aux sciences économiques et sociales. Il développe des recherches sur la prétopologie et ses applications.

Christine LARGERON
URA 934 Bt 101 Université Lyon I,
43 Bd du 11 Novembre 1918,
69100 VILLEURBANNE

Chercheur, membre de l'Unité Associée 934 du CNRS et de l'Université C. Bernard Lyon 1. Assistante en informatique à l'Université de St-Etienne. Prépare une thèse sur les problèmes de reconnaissance de formes en faisant appel à la géométrie informatique et les méthodes de relaxation.

RÉSUMÉ

Le papier que nous présentons propose une méthode de reconnaissance de formes que l'on pourrait situer parmi les techniques non paramétriques basées sur des procédures géométriques. Les idées qui y sont développées se retrouvent dans certaines approches de restauration d'images bruitées et dans des procédures d'apprentissage supervisé basées sur la relaxation.

Notre démarche s'articule autour de trois points :

* La définition d'une structure de voisinage possédant certaines propriétés.
* L'élaboration d'un critère d'homogénéité locale que l'on cherchera à optimiser en vue d'un réétiquetage.
* L'adoption d'une règle d'étiquetage pour des individus anonymes.

Nous concluerons par la présentation de quelques résultats expérimentaux.

MOTS CLÉ

Discrimination, Homogénéité locale, Réétiquetage, Voisinage.

SUMMARY

The paper we present here offers a method of pattern recognition which could be considered as a non-parametrical technique based on geometrical procedures. The ideas developed can be found in certain approaches for restoring pictures with sound and in supervised learning algorithms based on relaxation.

Our approach is based on three points :

* *The definition of a neighbourhood structure endowed with certain properties.*
* *Finding a local consistency criterion which we will try to optimize with a view to relabeling.*
* *Adopting a labeling rule for anonymous individuals.*

We will conclude by presenting a few experimental results.

KEY WORDS
Discrimination, Relabeling, Neighbourhood, Geometrical approach, local consistency, relaxation.

1. Introduction

Dans un article récent [17], M. Terrenoire et D. Tounissoux proposent une méthode de restauration d'images bruitées basée sur l'optimisation d'un critère d'homogénéité locale. D'autres auteurs ont également développé des approches voisines [7], [14]. Le présent papier qui en est inspiré tente d'adapter les idées qui sont développées dans [17] aux problèmes de classement appelés également de discrimination. En effet, la restauration d'une image bruitée, aussi bien à niveau de gris que binaire, s'apparente à un problème de discrimination et n'en constitue même qu'un aspect particulier. Sur une image binaire bruitée, par exemple, on cherche à déterminer l'état correct d'un pixel (0 ou 1). De même pour une image bruitée à niveau de gris, on cherche pour chaque pixel son niveau de gris réel.

En discrimination, il s'agit, pour chaque élément d'une population Π, de déterminer l'état d'une variable dite endogène. La particularité d'un problème de restauration d'images bruitées est que les points sont dans une trame qui est un sous-ensemble de $Z \times Z$ et que la variable endogène « niveau de gris » est de type ordinal. En revanche, en discrimination, l'espace de représentation est $\mathbb{R}^n$ et la variable endogène peut prendre des valeurs dans un ensemble d'étiquettes fini et non ordonné.

L'approche que nous préconisons est non paramétrique. De plus, elle ne cherche pas, comme en analyse discriminante ou dans certaines méthodes de régression, à établir un modèle mathématique reliant les variables endogène et exogènes. Elle s'apparente aux méthodes de relaxation et se veut essentiellement descriptive, fournissant néanmoins un procédure d'étiquetage pour un point non pris en compte dans la phase d'apprentissage.

Dans cet article, nous préciserons tout d'abord ce qu'est un problème de discrimination, ensuite nous développerons les 3 concepts de base sur lesquels s'articule notre démarche :

— le premier vise la définition d'une structure de voisinage dans $\mathbb{R}^n$. Au lieu des structures classiques telles que « les k-plus proches voisins » ou « le voisinage dans une hypersphère de rayon ε », nous préférons des voisinages non paramétriques basés sur des propriétés géométriques. De façon effective, nous utiliserons le voisinage défini selon le graphe de Gabriel [16]. Ce choix est justifié à la fois par des considérations mathématiques et algorithmiques ;

— le second a pour objet le choix d'une procédure de réétiquetage des points de l'ensemble d'apprentissage. Celle que nous préconisons est classique et est relativement simple. Elle consiste à réétiqueter un point suivant la majorité relative des étiquettes se trouvant dans son voisinage ;

— le troisième, qui constitue l'élément essentiel de notre contribution, est la proposition d'un critère dit « d'homogénéité locale » que nous chercherons à optimiser. En effet, nous considérons que des points voisins devraient avoir, dans leur voisinage, des étiquettes « semblables ». Cette ressemblance sera appréciée à partir de la notion de profil (obtenue par un recensement des étiquettes présentes dans le voisinage).

Nous terminerons cette partie méthodologique par une procédure d'identification des individus non pris en compte dans la phase d'apprentissage et par expérimentation sur un cas concret pris dans le domaine médical.

2. La discrimination

Soit Π une population d'individus ou d'objets concernée par le problème de reconnaissance de formes et Y une variable statistique à modalités discrètes. On supposera désormais que Y prend ses valeurs dans un ensemble E discret et de cardinal fini n. E est l'ensemble des étiquettes associées aux éléments de Π

$$Y : \Pi \to E = \{y_1, y_2, \ldots, y_n\}$$
$$\pi \to Y(\pi).$$

Par exemple, si Π est la population des étudiants et Y le résultat final de l'examen (succès : y_1, échec : y_2), alors $Y(\pi)$ sera le résultat de l'individu $\pi \in \Pi$ à l'examen en question (succès par exemple).

Dans la réalité, l'observation de $Y(\pi)$ pour tout $\pi \in \Pi$ n'est pas toujours facile et ceci pour des raisons diverses : délais trop longs, coût prohibitif, impossibilité, etc. Cela est très fréquent dans les sciences de l'homme. Par exemple, pour un malade atteint d'un nodule thyroïdien, le diagnostic (cancéreux ou bénin) nécessite dans de nombreux cas une intervention chirurgicale coûteuse, traumatisante et pas toujours indispensable pour le traitement du patient.

Dans un tel contexte, il sera utile de disposer d'une procédure de prévision φ c'est-à-dire une application de Π

dans E, grâce à laquelle nous pourrions pronostiquer l'état de la variable Y pour tous les individus $\pi \in \Pi$. Pour que cette règle de pronostic φ ait un intérêt, il faut qu'elle soit plus aisée à déterminer que Y et que, pour « une majorité de cas », on ait une bonne prédiction c'est-à-dire :

$$\varphi(\pi) = Y(\pi) \quad \pi \in \Pi .$$

La détermination de φ est liée à l'hypothèse selon laquelle les valeurs prises par la variable statistique Y ne relèvent pas du hasard mais de certaines situations particulières que l'on peut caractériser. Pour cela, l'expert du domaine concerné (médecin par exemple) établit une liste a priori de variables statistiques appelées variables exogènes et notées

$$X_1, ..., X_j, ..., X_p$$

où X_j est une application de Π dans un ensemble $\mathscr{R}_j$.

On note :

$$X : \Pi \to \mathscr{R} = \mathscr{R}_1 \times \cdots \times \mathscr{R}_j \times \cdots \times \mathscr{R}_p$$
$$\pi \to X(\pi) = (X_1(\pi), ..., X_j(\pi), ..., X_p(\pi))$$

$\mathscr{R}$ est appelé « Espace de représentation ». Si toutes les variables X_j ($j = 1, ..., p$) prennent leur valeur sur $\mathbb{R}$, alors $\mathscr{R} = \mathbb{R}^p$.

La construction de φ passe généralement par 2 étapes.

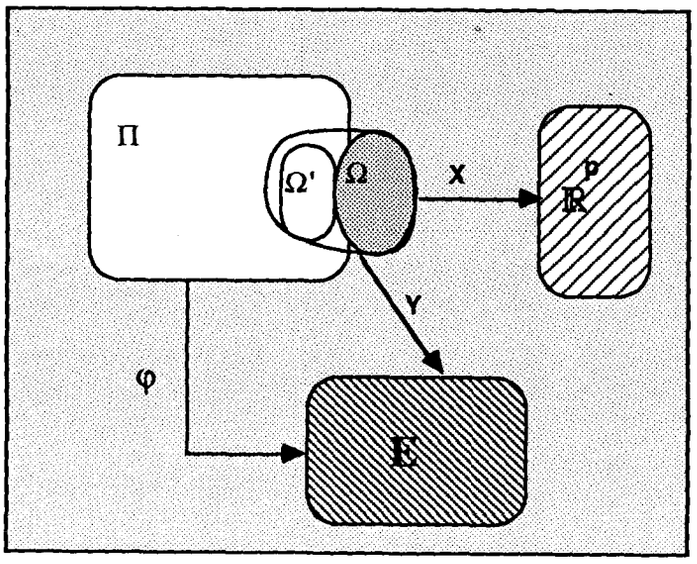
L'apprentissage :

Dans cette étape, on prélève de la population Π concernée par le problème de la discrimination deux échantillons Ω et Ω'. Le premier dit « d'apprentissage » servira à la construction de φ et le second dit « de test », servira à tester la validité de φ.

L'étiquetage :

Dans cette étape, si notre règle de prédiction φ est jugée satisfaisante, c'est-à-dire si pour une majorité d'éléments π de l'échantillon test Ω' on a $\varphi(\pi) = Y(\pi)$, nous pourrons l'utiliser pour prévoir $Y(\pi)$ sur tous les éléments anonymes qui ne sont pas dans les échantillons Ω et Ω', c'est-à-dire $\pi \in \Pi - (\Omega \cup \Omega')$.

D'où le schéma :



3. Les données

Le problème de la discrimination étant posé, voyons comment, concrètement, va se dérouler le processus de construction.

Soit ω un individu de l'échantillon d'apprentissage Ω. La phase d'apprentissage exige que l'on dispose pour tout $\omega \in \Omega$:

1° Des p valeurs prises par les p variables statistiques $X_1, ..., X_j, ..., X_p$ appelées variables exogènes ou explicatives qu'on supposera quantitatives :

$$X(\omega) = (X_1(\omega), ..., X_j(\omega), ..., X_p(\omega)) \in \mathbb{R}^p .$$

2° De l'état de la variable endogène Y dite à expliquer :

$$Y(\omega) \in E = \{y_1, ..., y_n\} .$$

E est appelé ensemble des étiquettes.

Lors de la phase d'étiquetage, on se contentera de connaître seulement $X(\pi)$ pour $\pi \in \Pi$.

4. Notions de base

4.1. *p*-VOISINAGE (1)

Soit pour tout $(\omega, \omega') \in \Omega \times \Pi$ une proposition relative à ω et ω' notée $p(\omega, \omega')$. Pour tout $\omega' \in \Pi$, on note $V(\omega')$ l'ensemble des points $\omega \in \Omega$ vérifiant la propriété p :

$$V(\omega') = \{\omega \in \Omega / p(\omega, \omega') \text{ vraie}\}$$

et on l'appelle p-voisinage de ω'.

Pour la construction de la structure de ce p-voisinage, de multiples choix sont possibles : les k-plus proches voisins [1], [16], les points se trouvant dans une boule de rayon ε (ε-voisin) [17], le voisinage de Delaunay [2], [20], [21], la sphère d'influence, la « lunule » [18], [19], etc.

Le critère d'homogénéité locale que nous construirons au § 5 et les calculs auxquels il conduit pour son optimisation supposent que notre structure de voisinage soit symétrique, c'est-à-dire que pour $\omega \in \Omega$ si ω est p-voisin de ω' alors ω' est p-voisin de ω. Cette exigence exclut les k-plus proches voisins, les ε-voisins et tous ceux qui ne possèdent pas cette propriété de symétrie.

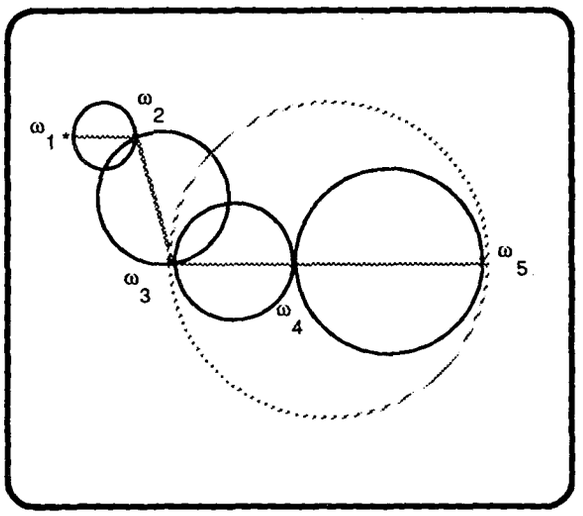
Dans ce papier, nous retenons le voisinage selon le graphe de Gabriel [16], [19] car il vérifie la propriété de symétrie et, de plus, sa construction est relativement simple à programmer comparativement au voisinage associable à la triangulation de Delaunay par exemple.

(1) Le mot « voisinage » est utilisé ici de façon abusive car la définition que nous en donnons n'est pas en conformité avec les définitions de la topologie, d'où l'appellation de p-voisinage.

La propriété p définissant le voisinage selon le graphe de Gabriel peut sénoncer comme suit [16] :

$p(\omega, \omega')$ est vraie si et seulement si le disque ouvert de diamètre $\|X(\omega) - X(\omega')\|$ ne contient aucun autre point de Ω.

Considérons l'exemple de 5 points placés dans $\mathbb{R}^2$ comme le montre la figure suivante :



Sur cette figure

$$V(\omega_1) = \{\omega_2\}$$
$$V(\omega_2) = \{\omega_1, \omega_3\}$$
$$V(\omega_3) = \{\omega_2, \omega_4\}$$
$$V(\omega_4) = \{\omega_3, \omega_5\}$$
$$V(\omega_5) = \{\omega_4\}$$

ω_3 n'appartient pas à $V(\omega_5)$ car le disque de diamètre $\|X(\omega_3) - X(\omega_5)\|$ contient le point ω_4.

4.2. PROFIL

DÉFINITION : *Soit*

$$S_n = \{(\gamma_0, \gamma_1, ..., \gamma_n) \in \mathbb{R}^{n+1} ;$$
$$\forall i = 0, ..., n, \quad \gamma_i \geq 0 \quad \text{et} \quad \Sigma\gamma_i = 1\}$$

le simplexe de dimension $(n+1)$, *où* n *est le cardinal de* E. *On appelle profil sur* Ω *toute application* ℓ *de* Ω *dans* S_n. *Si* $\omega \in \Omega$, $\ell(\omega)$ *sera un profil de* ω. *On notera* $\mathcal{L}(\Omega)$ *l'ensemble des profils.*

Dans ce qui suit, on se propose d'utiliser un profil qui permette de caractériser l'environnement de ω, c'est-à-dire les points de $V(\omega)$, au moyen de la proportion de points de $V(\omega)$ étiquetés y_k. Pour cela, nous construisons le profil ℓ de la façon suivante : on adjoint à $E = \{y_1, ..., y_n\}$ un élément $y_0 \notin E$ qui sera une nouvelle modalité de Y, attribuée aux points pour lesquels aucune décision d'affectation ne sera possible. Cette étiquette y_0 n'interviendra que lors du réétiquetage de tous les $\omega \in \Omega$.

Soit pour $k = 0, ..., n$,

$$C_k = Y^{-1}(\{y_k\}) = \{\omega \in \Omega / Y(\omega) = y_k\}.$$

C_0 est donc le groupe des points non étiquetés. Comme au départ, tous les points sont déjà étiquetés, on a $C_0 = \varnothing$ et pour $k \neq 0$ on a $C_k \neq \varnothing$.

Lors du réétiquetage, d'une part certains points conserveront la même étiquette (on espère que cela soit ainsi pour le plus grand nombre) et d'autre part C_0 peut rester vide (ce qui est souhaitable).

L'application $\ell : \Omega \to \mathbb{R}^{n+1}$ définie par

$$\ell(\omega) = (\ell_0(\omega), \ell_1(\omega), ..., \ell_k(\omega), ..., \ell_n(\omega)) \in \mathbb{R}^{n+1}$$

avec

$$\forall k = 0, ..., n \quad \ell_k(\omega) = \frac{\text{card}\,(V(\omega) \cap C_k)}{\text{card}\,(V(\omega))}$$

est ainsi un profil qui à tout $\omega \in \Pi$, fait correspondre la proportion de points p-voisins à ω qui figurent dans C_0, C_1, ..., C_j, ..., C_n.

4.3. RÉÉTIQUETAGE

Le réétiquetage (ne concerne que les individus de l'échantillon d'apprentissage) d'un point $\omega \in \Omega$ doit, comme nous l'avons précisé au § 4.2, tenir compte de l'environnement de ω qui est résumé par la notion de profil.

DÉFINITION : *Nous appelons réétiquetage toute application* ε *de* S_n *dans* $E^* = \{y_0, ..., y_n\}$.

Étant donné un réétiquetage ε, l'application $\varepsilon \circ \ell$ notée simplement $\varepsilon\ell$ est une application de Ω dans E^* qui, à tout $\omega \in \Omega$, permet d'associer un élément de E^* appelé étiquette de ω.

Nous souhaiterions évidemment que $\varepsilon\ell$ soit telle que $\varepsilon\ell(\omega) = Y(\omega)$ pour tout ω : cela prouverait que l'on peut reconstituer Y en combinant ε et ℓ. Si l'on pouvait alors étendre ℓ à Π tout entier, on disposerait ainsi d'un moyen de prédire ce que pourrait être $Y(\omega')$ pour tout $\omega' \in \Pi - \Omega$.

Il va de soi que le choix de ε devra être fait en fonction de celui de ℓ. Compte tenu du choix fait au § 4.2, il semble naturel de prendre pour ε l'une ou l'autre des applications ε_1 ou ε_2 ci-après.

a) Pour $\gamma \in S_n$, on prend :

* $\varepsilon_1(\gamma) = y_k$ s'il existe $k \in \{1, ..., n\}$ tel que $\forall i \neq k$, $\gamma_i < \gamma_k$.
* $\varepsilon_1(\gamma) = y_0$ sinon.

Ainsi, ω est étiqueté y_k si et seulement si $\ell_k(\omega)$ est la plus grande des composantes de $\ell(\omega)$ (c'est la règle du vote à la majorité relative).

b) On imagine une procédure plus exigeante qui refuserait à ω d'être étiqueté y_k même si $\ell_k(\omega)$ est la plus grande des composantes de $\ell(\omega)$, lorsque cette composante ne paraît pas suffisamment élevée. On choisit alors un seuil $s \in]0, 1[$ et on définit ε_2 en prenant, pour $\gamma \in S_n$:

* $\varepsilon_2(\gamma) = y_k$ si $\forall i \neq k$; $i, k \in \{1, ..., n\}$, $\gamma_i < \gamma_k$ et $s \leq \gamma_k$

* $\varepsilon_2(\gamma) = y_0$ si $\max_{i=1,\ldots,n} \{\gamma_i\} < s$ ou bien le max n'est pas unique.

Ainsi, avec l'un de ces choix, tout $\omega \in \Omega$ se voit attribuer l'étiquette y_k si les points de C_k sont, en majorité, dans l'ensemble des points p-voisins à ω. D'autres règles de réétiquetage plus élaborées sont possibles notamment celles basées sur le principe de la relaxation : [4], [9], [13].

Il n'y a malheureusement aucune raison pour que l'on ait $Y(\omega) = \varepsilon \ell(\omega)$ pour tout $\omega \in \Omega$, et de ce fait, notre procédure de reconstitution de Y n'est certainement pas parfaite ; on pourrait chercher à l'améliorer soit en modifiant ℓ, soit en modifiant ε, soit en modifiant les deux.

Nous proposons ci-dessous une procédure visant simplement à modifier ℓ en optimisant un critère.

5. Algorithme de réétiquetage basé sur un critère d'homogénéité locale

Soit ℓ le profil déjà construit au § 4.2 et soit $\ell' \in \mathscr{L}(\Omega)$ un autre profil que nous allons lui substituer. Le principe de la procédure de modification du profil ℓ est articulé autour de deux idées :

***a)* Notion de compatibilité :** le profil ℓ est construit à partir de l'observation directe de Y, il serait donc souhaitable que le profil ℓ' que nous allons lui substituer ne soit pas « trop éloigné de ℓ ».

***b)* Notion de cohérence :** les points p-voisins de $\omega \in \Omega$ sont en fait des points présentant vis-à-vis des variables exogènes $X_1, \ldots, X_j, \ldots, X_p$ des caractéristiques voisines (cela provient du choix de la propriété p cf. § 4.1) ; comme nous espérons prédire l'état de la variable endogène à partir des variables exogènes, nous ne pourrons y parvenir que si des caractéristiques exogènes proches se traduisent généralement par des valeurs identiques de l'endogène. Il s'agit là d'une hypothèse que nous supposons réalisée. Dans ces conditions il faudra que pour tout $\omega \in \Omega$, les points p-voisins à ω aient un profil le plus proche possible de celui de ω.

D'où :

DÉFINITION : *Nous appellerons profil ℓ-efficace tout profil $\ell' \in \mathscr{L}(\Omega)$, tel que*

$$\delta(\ell') = \sum_{\omega \in \Omega} \left[\| \ell(\omega) - \ell'(\omega) \|^2 + \frac{1}{\text{card } (V(\omega))} \times \times \sum_{\omega' \in V(\omega)} \| \ell'(\omega) - \ell'(\omega') \|^2 \right]$$

soit minimum.

Le premier terme de l'expression ci-dessus traduit une sorte de compatibilité entre le profil observé ℓ et le profil calculé ℓ'. Le second terme, quant à lui, traduit une notion de cohérence locale qui n'existait pas nécessairement avant et que l'on souhaite établir par ce calcul. Nous cherchons les vecteurs $\ell'(.)$ qui rendraient minimum $\delta(\ell')$ qui représente ce que nous appelons une « homogénéité locale ».

Ce problème peut être facilement résolu par une technique d'optimisation. En effet, l'expression ci-dessus peut s'écrire comme suit :

$$\delta(\ell') = \sum_{k=1}^{n} \left[\sum_{\omega \in \Omega} \left((\ell_k(\omega) - \ell'_k(\omega))^2 + \frac{1}{\text{card } (V(\omega))} \times \times \sum_{\omega' \in V(\omega)} (\ell'_k(\omega) - \ell'_k(\omega'))^2 \right) \right]$$

que l'on notera :

$$\delta(\ell') = \sum_{k=1}^{n} \delta(\ell'_k) .$$

Puisque les étiquettes sont indépendantes, on optimisera suivant chacun des $\delta(\ell'_k)$, $k = 1, \ldots, n$. En réécrivant l'équation précédente pour un $\omega \in \Omega$, on aura à chercher l'optimum de l'expression ci-dessous :

$$\delta(\ell'_k) = \sum_{\omega \in \Omega} \left[(\ell_k(\omega) - \ell'_k(\omega))^2 + \frac{1}{\text{card } (V(\omega))} \times \times \sum_{\omega_j \in V(\omega)} (\ell'_k(\omega) - \ell'_k(\omega_j))^2 + \sum_{\omega_i \in V(\omega)} \frac{1}{\text{card } (V(\omega_i))} (\ell'_k(\omega) - \ell'_k(\omega_i))^2 \right] .$$

Cette dernière écriture n'est possible qu'à condition que pour tout $\omega_i \in \Omega$ et tout $\omega_j \in \Omega$, si $\omega_i \in V(\omega_j)$ alors $\omega_j \in V(\omega_i)$. Cette condition est vérifiée dans notre cas car la structure de voisinage que nous avons retenue (p-voisinage selon le graphe de Gabriel cf. § 4.1) possède cette propriété de symétrie.

La convexité stricte de $\delta(\ell'_k)$ se démontre sans difficulté. $\ell'(.)$ est un point de $\mathbb{R}^{\text{card } (\Omega)}$.

Le minimum sera un point de gradient nul. On aura donc à déterminer pour tout ω, $\ell'_k(\omega)$ pour $k = 1, \ldots, n$. Soit :

$$\frac{\partial \delta}{\partial \delta(\ell'_k(\omega))} = 2\ell'_k(\omega) - 2\ell_k(\omega) + 2\ell'_k(\omega) \times \times \left[1 + \sum_{\omega_i \in V(\omega)} \frac{1}{\text{card } (V(\omega_i))} \right] - 2 \sum_{\omega_i \in V(\omega)} \ell'_k(\omega_i) \left[\frac{1}{\text{card } (V(\omega))} + \frac{1}{\text{card } (V(\omega_i))} \right] .$$

Ainsi, pour tout $\omega \in \Omega$, et $k = 1, \ldots, n$ on aura :

$$\ell'_k(\omega) = \frac{\ell_k(\omega) + \sum_{\omega_i \in V(\omega)} \ell'_k(\omega_i) \left[\frac{1}{\text{card } (V(\omega))} + \frac{1}{\text{card } (V(\omega_i))} \right]}{\left[2 + \sum_{\omega_i \in V(\omega)} \frac{1}{\text{card } (V(\omega_i))} \right]} .$$

Cette équation de point fixe peut être résolue efficacement par la méthode de Gauss-Seidel [6]. Nous avons démontré que les conditions suffisantes de convergence de l'algorithme de Gauss-Seidel sont vérifiées. Ainsi la méthode itérative est convergente quel que soit le point de départ. En pratique nous choisissons pour ℓ^0 le profil ℓ que nous avons observé sur l'échantillon d'apprentissage, d'où l'itération générale suivante :

$$\ell_k^{\prime(j+1)}(\omega) = \frac{\ell_k(\omega) + \sum_{\omega_i \in V(\omega)} \ell_k^{\prime(j)}(\omega_i)\left[\frac{1}{\text{card}\,(V(\omega))} + \frac{1}{\text{card}\,(V(\omega_i))}\right]}{\left[2 + \sum_{\omega_i \in V(\omega)} \frac{1}{\text{card}\,(V(\omega_i))}\right]}$$

On montre que pour tout $\omega \in \Omega$ on a $\sum_{k=1}^{n} \ell_k'(\omega) = 1$ et $\ell_k'(\omega) \geq 0$, pour $k = 1, ..., n$, c'est la raison pour laquelle nous n'avons pas introduit ces contraintes dans le programme d'optimisation...

Pour le réétiquetage des points de l'échantillon Ω, on procèdera comme suit :

disposant, pour chaque point $\omega \in \Omega$, d'un profil ℓ-efficace $\ell'(\omega)$, nous réétiquetons l'ensemble des points Ω suivant l'une des deux procédures décrites au § 4.3. Pour les exemples que nous traiterons c'est la seconde règle que nous avons retenue, c'est-à-dire qu'un point $\omega_i \in \Omega$ sera réétiqueté y_k si la composante $\ell_k'(\omega)$ est la plus grande et que, de plus, elle est supérieure à un seuil s.

6. Procédure d'extension

Cette procédure d'extension est indispensable si on veut disposer d'une règle d'affectation pour un individu test ou anonyme.

Disposant d'un profil ℓ-efficace $\ell'(.)$, nous lui associons une application $\hat{\ell}'$ de Π dans S_n définie ainsi :

1° Pour tout $k = 0, ..., n$ on note

$$C_k' = \{\omega \in \Omega / \varepsilon\ell'(\omega) = y_k\} .$$

2° Pour tout $\omega \in \Pi$ on pose :

$$\hat{\ell}_k'(\omega) = \frac{\text{card}\,(V(\omega) \cap C_k')}{\text{card}\,(V(\omega))}$$

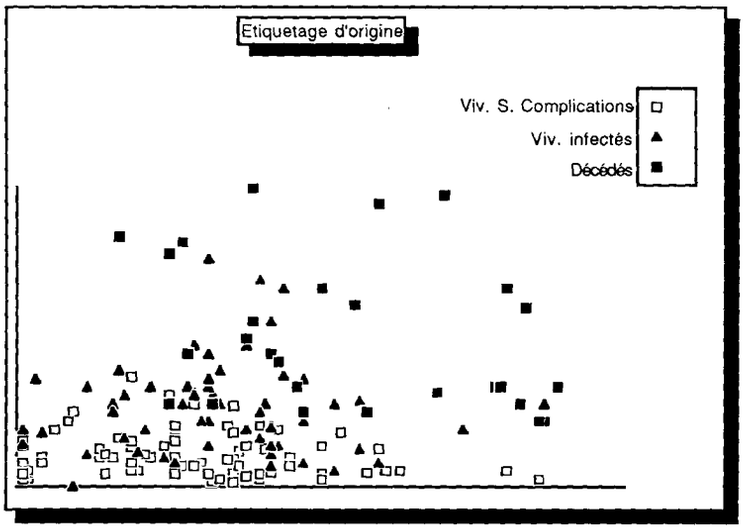
et $$\hat{\ell}'(\omega) = (\hat{\ell}_0'(\omega), \hat{\ell}_1'(\omega), ..., \hat{\ell}_n'(\omega)) .$$

Il est certain que $\hat{\ell}'(\omega) \in S_n$ et on dira que $\hat{\ell}'$ est le profil étendu associé à ℓ'. Ainsi, pour $\omega \in \Pi - \Omega$, on prend pour prédiction de Y : $\varepsilon\hat{\ell}'(\omega)$.

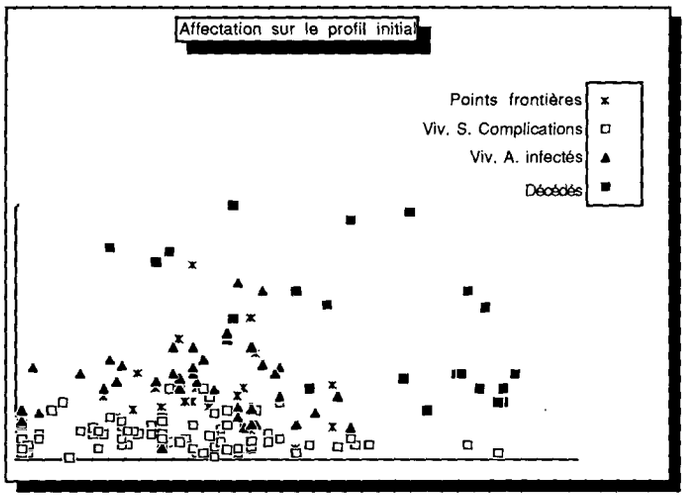
Nous pourrons ainsi, quand un individu anonyme se présente, déterminer ses voisins dans Ω, suivant les étiquettes de ceux-ci (après réétiquetage) et suivant la règle de réétiquetage retenue, pronostiquer son étiquette inconnue.

7. Application

L'exemple que nous présentons est issu d'une application médicale [3], [15]. Il s'agit de mettre au point une procédure d'évaluation du risque infectieux chez le grand brûlé. Trois groupes (vivants sans complications, vivants avec une complication septique, décédés à la suite d'un choc septique) sont à distinguer. 174 patients formant l'échantillon d'apprentissage ont été retenus. La figure ci-dessous représente la répartition dans 3 groupes (87 vivants sans complications, 60 vivants avec une septicémie, 27 décédés à la suite d'une infection). Les axes des abscisses et des ordonnées représentent l'âge et la surface de brûlure du patient.



En se basant simplement sur les profils initiaux $\ell(\omega)$, $\omega \in \Omega$, le réétiquetage des individus de l'échantillon d'apprentissage Ω donne les résultats suivants :



Etiquetage original	Réétiquetage (affectation)			
	Vivant Sans complications	Vivant infectés	Décédés	Non affectés. Pts. Frontière
Vivant Sans complications	71	11	0	5
Vivant infectés	15	27	3	15
Décédés	0	6	18	3

Proportion de points bien affectés : 66,6 %.

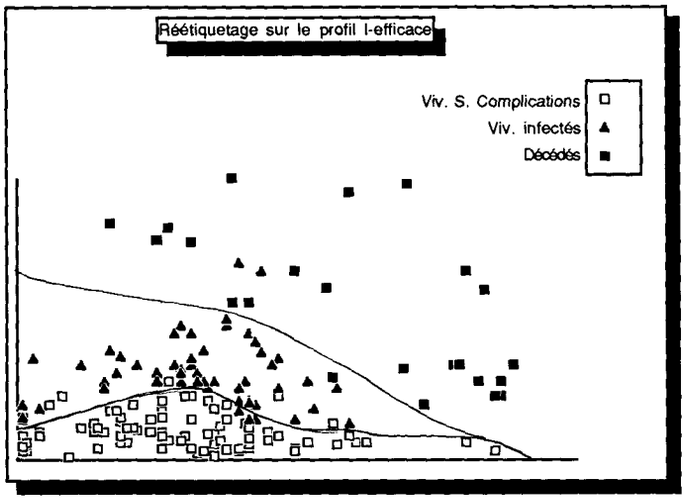
On considère qu'un point est bien affecté s'il n'a pas changé d'étiquette après le réétiquetage.

Le calcul du profil ℓ-efficace et le réétiquetage de chaque élément ω de l'échantillon d'apprentissage Ω en se basant sur le nouveau profil donnent un taux de reconnaissance sensiblement meilleur (proportion d'individus qui n'ont pas changé d'étiquette après réétiquetage) de 73,5 %.

Etiquetage original	Réétiquetage (affectation)			
	Vivant Sans complications	Vivant infectés	Décédés	Non affectés. Pts. Frontière
Vivant Sans complications	73	14	0	0
Vivant infectés	20	35	5	0
Décédés	0	7	20	0

Proportion de points bien affectés : 73,5 %.

La figure ci-dessous représente le résultat du réétiquetage obtenu sur l'échantillon d'apprentissage. Notons que



l'analyse discriminante a donné un taux de reconnaissance de 72,4 %. D'autres techniques non paramétriques [3] n'ont pas donné des résultats plus significatifs.

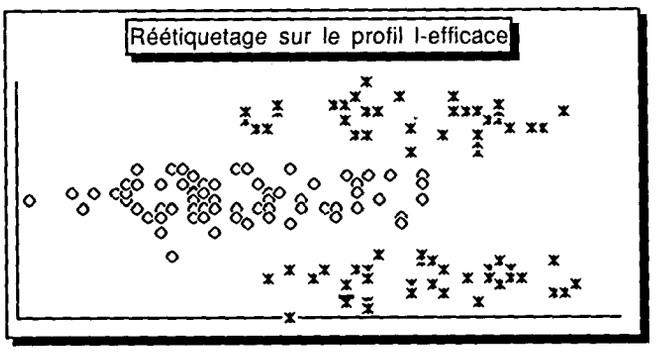
Afin de mieux apprécier la qualité des résultats fournis par notre méthode, il semble judicieux d'affiner l'expérimentation. Pour cela nous avons adopté le protocole suivant :

les 174 individus de l'échantillon d'apprentissage ont été ventilés au hasard sur 5 paquets en respectant les proportions dans chacun des 3 groupes à discriminer. Nous avons appliqué 5 fois notre algorithme en prenant à tour de rôle un paquet comme échantillon test, les autres étant réunis pour constituer l'ensemble d'apprentissage. Les résultats obtenus sont consignés dans le tableau suivant. Nous donnons également les résultats obtenus par l'analyse discriminante sur les mêmes données.

Essai N°	Effectifs		Proportion de biens classés			
			Notre approche		Analyse discriminante	
	Base	Test	Base	Test	Base	Test
1	139	35	75,54%	71,43%	71,74%	71,43%
2	139	35	72,66%	77,14%	71,22%	74,29%
3	140	34	75,71%	67,65%	74,29%	70,59%
4	139	35	75,54%	63,00%	72,66%	74,29%
5	139	35	74,10%	68,57%	56,12%	68,57%
Taux moyen de reconnaissance :			74,71%	69,55%	69,20%	71,83%

Il ressort, en effet, une stabilité dans les résultats. L'un des avantages de l'algorithme proposé est que, sur une configuration où les groupes ne sont pas linéairement séparables, les résultats demeurent bons alors qu'ils se détériorent avec l'analyse discriminante.

L'exemple, ci-dessous, issu de [5], permet de constater une reconnaissance presque parfaite.



Les résultats sont également très satisfaisants quand les individus sont placés dans un espace à plus de 2 dimensions. L'utilisation des données extraites de [11] relatives à la discrimination entre chiens et loups donne un taux de reconnaissance de 93 %. Dans cet exemple, 5 mesures ont été relevées sur 43 crânes dont 30 de chiens et 13 de loups.

8. Conclusion

L'idée de fond de ce papier est basée sur le principe qui dit « qui s'assemblent se ressemblent » ce qui se traduit dans notre langage par « des points voisins dans l'espace de représentation devraient avoir la même étiquette ».

Dans un premier temps, il fallait donner un sens précis au terme de voisins. Nous avons examiné de multiples possibilités et nous distinguons, en fait, deux grandes classes de structures de voisinage :

1° Celle qui exige la fixation d'un paramètre par l'utilisateur. Par exemple les k-plus proches voisins, la boule de rayon ε... L'inconvénient majeur de ces structures de voisinage est que le résultat de l'apprentissage est lié au bon choix des paramètres k, ε, ... Leur avantage est qu'elles peuvent conduire à des approches probabilistes [10], [13].

2° Les structures de voisinage que nous avons préférées ne sont pas paramétrées, elles font appel à des propriétés géométriques liées à l'espace de représentation. Par exemple, les voisinages selon Voronoi, Delaunay, le graphe de Gabriel, etc. Ces structures sont connues et ont été souvent utilisées en reconnaissance de formes [19]. L'inconvénient de celles-ci est la complexité de leur algorithme quand on se place dans $\mathbb{R}^P$ [16].

Dans un second temps, on observe si les individus voisins ont « en général » même étiquette. Comme ce n'est souvent pas le cas, on se propose alors de réétiqueter l'ensemble des individus afin de créer artificiellement cette situation. Là aussi de nombreuses approches sont possibles, par exemple le réétiquetage suivant la majorité dans le voisinage ou bien par relaxation. Nous proposons un critère permettant le réétiquetage qui tient compte de la structure locale de chaque point. En ce sens, notre procédure s'apparente aux méthodes de relaxation.

Nous avons expliqué au § 4.1 les raisons du choix du voisinage selon le graphe de Gabriel. Précisons cependant que ce voisinage peut s'avérer inadapté dans certaines configurations. Par exemple, pour discriminer deux groupes situés le long de deux droites parallèles, l'analyse discriminante peut fournir de meilleurs résultats.

En ce qui concerne le critère d'homogénéité d'autres choix sont actuellement à l'étude : ils reposent sur la prise en compte non seulement des voisins immédiats de chacun des points mais aussi des voisins au k-ième rang en pondérant le cas échéant « la compatibilité » ou « la cohérence locale »... ce qui ouvre la voie à une formalisation statistique.

Pour ce qui est de l'expérimentation, nous réalisons actuellement de nombreux essais, réels ou simulés, comparatifs avec de nombreuses méthodes : relaxation, k-plus proches voisins, ε-voisins, discriminante, ... Si les résultats expérimentaux que nous avons, jusqu'à présent, obtenu ne sont pas, systématiquement, sensiblement meilleurs que ceux des autres méthodes citées précédemment, les mieux adaptées au problème que l'on traite, il n'en demeure pas moins que l'approche que nous proposons présente des avantages pratiques. Par exemple, l'utilisateur n'a pas à rechercher manuellement le bon paramètre k ou ε, ... ni la procédure la mieux appropriée. De plus, notre méthode s'adapte à la structure géométrique des formes que l'on cherche à discriminer et présente une faible complexité algorithmique aussi bien dans la recherche des voisinages que dans l'optimisation du critère d'homogénéité locale.

Manuscrit reçu le 19 juin 1989.

BIBLIOGRAPHIE

[1] P. Devijver, *Selection of prototypes for nearest neighbour classification.* Indian statistical institute golden jubilee. Proc. Int. Conf. on advances in information sciences and technology, 1982.

[2] P. Devijver & M. Dekesel, *Computing multidimensional Delaunay tesselation.* Report R.464 Philips Research Laboratory, Brussels, 1982.

[3] R. Fages, M. Terrenoire, D. Tounissoux, A. Zighed, *Non supervised classification tools adapted for supervised classification.* Nato ASI Series Vol. F30 Springer-Verlag, 1985.

[4] O. D. Faugeras & M. Berthod, Improving consistency and reduction ambiguity in stochastic labelling : an optimization approach. IEEE Vol. PAMI-30, No. 4, July 1981.

[5] K. Fukunaga, *Introduction to statistical pattern recognition.* Academic Press, 1972.

[6] R. Gastinel, *Analyse numérique linéaire.* Herman Paris 1966.

[7] X. Guyon & J. F. Yao, *Analyse discriminante contextuelle.* Actes 5e Jr. Anl. des données et informatique. Tome 1, p. 43-52, 1987.

[8] Hossam A. El Gindy, Godfried T. Toussaint, *Computing the relative neighbour decomposition of simple polygon.* Computational morphology. Ed. G. T. Toussaint, North-Holland, 1988.

[9] R. A. Hummel & S. W. Zucker, *On the foundation of relaxation labeling.* IEEE vol. PAMI-5, No. 3, May 1983.

[10] J. Illingworth & J. Kittler, *Optimisation algorithms in probabilistic relaxation labelling Pattern recognition Theory and application.* Nato series Vol. 30 Springer Verlag, 1987.

[11] M. Jambu, M. O. Lebeaux, *Classification automatique pour l'analyse des données.* Dunod, 1978.

[12] A. K. Jain, *Advances in statistical pattern recognition, Pattern recognition Theory and application.* Nato series Vol. 30 Springer Verlag, 1987.

[13] J. Kittler, *Relaxation labelling, Pattern recognition Theory and application.* Nato series Vol. 30 Springer Verlag, 1987.

[14] M. Levy, *A new theoriktical approach to relaxation, application to edge detection.* Actes IAPR Rome, 1988, IEEE.

[15] J. Marichy, G. Buffet, A. Zighed, P. Laurent, *Early detection of septicemia in burnt patients,* Actes 3rd Int. Conf. Sci. in Health Care, p. 505-508, Munich 1984. Ed. Springer-Verlag.

[16] F. P. Preparata & M. I. Shamos, *Computational geometry : an introduction.* Springer-Verlag, 1988.

[17] M. Terrenoire, D. Tounissoux, *Restauration d'image par optimisation d'un critère d'homogénéité locale.* Proceding of workshop on syntactical and structural pattern recognition, Pont-à-Mousson, 1988.

[18] G. T. Toussaint, *Computational geometry recent relevant to pattern recognition.* Nato Asi Series Vol. F30, Springer-Verlag, 1987.

[19] G. T. Toussaint, *A Graph-theoretical primal sketch.* Computational morphology, Ed. Toussaint, North-Holland, 1988.

[20] J. I. Toriwaki, S. Yokoi, *Voronoi andrelated neighbors on digitized 2-dimensiona space with application of texture analysis.* Computational morphology. Ed. Toussaint, North-Holland, 1988.

[21] D. F. Watson, *Computational the n-dimensional Delaunay tesselation with application to voronoi polytopes.* The computer journal. Vol. 24, No. 2, 1981.