archives-ouvertes.fr

Réduction de dimension par un nouvel estimateur de la distance de patrick Fisher é l'aide des fonctions orthogonales

Faouzi Ghorbel, Wissal Drira

To cite this version:

Faouzi Ghorbel, Wissal Drira. Réduction de dimension par un nouvel estimateur de la distance de patrick Fisher é l'aide des fonctions orthogonales. 42èmes Journées de Statistique, 2010, Marseille, France, France. inria-00494761

HAL Id: inria-00494761
https://hal.inria.fr/inria-00494761

Submitted on 24 Jun 2010

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

RÉDUCTION DE DIMENSION PAR UN NOUVEL ESTIMATEUR DE LA DISTANCE DE PATRICK FISHER À L'AIDE DES FONCTIONS ORTHOGONALES

Wissal DRIRA & Faouzi GHORBEL

Laboratoire CRISTAL, pôle GRIFT, Ecole National des Sciences de l'Informatique,
Campus universitaire de La Manouba, 2010 Manouba, Tunisie
wissal.drira@gmail.com , faouzi.ghorbel@ensi.rnu.tn

Abstract

Here, we intend to introduce new estimators of the Patrick Fisher distance and the Euclidian probabilistic dependence measure by using orthogonal series. Therefore non parametric multiclass dimension reduction could be easily generalized to the multivariate case reduction. The performance of the proposed estimates will be studied by simulations in the case of mixtures of non Gaussian multivariate distributions. Such estimators will be applied in Face recognition.

Résumé

Dans cet article, nous proposons d'introduire deux nouveaux estimateurs. Un premier pour la distance de Patrick Fisher qui permet la sélection des primitives dans le cas de deux classes. Le deuxième estime la mesure de dépendance probabiliste dans le cas multi classes. Ces estimateurs seront appliqués dans le cadre de la reconnaissance faciale des visages.

1. Introduction

Un des problèmes les plus étudiés et essentiels dans le domaine de la reconnaissance de formes statistique est la sélection de primitives discriminantes. En reconnaissance faciale des visages, en reconnaissance de caractères manuscrits et en indexation de bases d'images par le contenu, le nombre de descripteurs de l'image ou de la forme est souvent très grand. Afin de mettre en œuvre une classification de ces images à partir de leurs descripteurs, une modélisation probabiliste est souvent requise. L'ensemble des descripteurs est souvent représenté mathématiquement dans le cas le plus simple, par un vecteur aléatoire. Sa loi de probabilité appelée loi mélange et les lois relatives à une classe, connues sous le nom de lois conditionnelles, souvent supposées absolument continues, doivent être estimées pour pouvoir appliquer une règle de classification donnée telle que la règle de Bayès ou des plus proches voisins. D'autre part, il est bien connu que l'estimation des lois des vecteurs aléatoires en grande dimension nécessite des tailles d'échantillon très grandes voire pas réalistes. En effet, pour une précision donnée, les théorèmes de convergence de l'histogramme, de la méthode du noyau, des fonctions orthogonales ou de l'histospline, tous des estimateurs des densités de probabilité, recquièrent des tailles d'échantillons qui augmentent exponentiellement avec la dimension D du vecteur aléatoire des primitives. L'idée de la réduction de dimension est réaliste car l'estimation des critères de séparabilité à optimiser, est réalisée dans l'espace réduit de dimension $d << D$. Cependant, lorsque le nuage représentant les observations relatives à une classe, occupent un domaine de l'espace de forme non connexe, la réduction de dimension par la méthode des matrices de dispersion donne des résultats non satisfaisants (voir Figure 1). Patrick et Fisher [2] ont proposé une solution non paramétrique basée sur des distances entre les densités de probabilité conditionnelles. Ils proposent d'utiliser l'estimateur à noyau pour la distance euclidienne entre des fonctions de carré intégrable. Dans un contexte de classification d'image à partir de paramètres sur les textures, Alain Hillion et al proposent une sélection de primitives scalaire. Au sens du minimum de la probabilité d'erreur, ils montrent la meilleure tenue de l'analyse discriminante obtenue par la distance de Patrick Fisher vis-à-vis de celle se basant sur les matrices de dispersion. Toutefois, l'extension d'une telle méthode à la réduction vectorielle $(d > 1)$ ne semble pas être aisée. Des méthodes itératives ou récursives ont été proposées dans la littérature [2]. Dans cet article, nous proposons d'introduire de nouveaux estimateurs pour la distance de Patrick Fisher et de la mesure de dépendance probabiliste en se basant sur les fonctions orthogonales déjà utilisées pour l'estimation des densités de probabilité. Une telle méthode permet l'extraction scalaire et peut être aisément généralisée au cas de la réduction multidimensionnelle. Ainsi le présent article sera organisé comme suit. Dans un premier paragraphe, nous rappelons la formulation de l'analyse discriminante qui se base sur les matrices de dispersion et celle qui optimise l'estimateur à noyau de la distance de Patrick Fisher. Les estimateurs proposés de la distance et de la mesure de dépendance probabiliste qui sont exprimés dans un espace de Hilbert muni d'une base de fonctions orthogonales seront introduit dans la seconde section. Dans le troisième paragraphe des études asymptotiques seront présentées. Quelques simulations et résultats expérimentaux en reconnaissance faciales illustreront l'intérêt pratique des ces nouveaux estimateurs.

2. Analyse discriminante basée sur les matrices de dispersion

Une des méthodes de réduction les plus connus est celle de Fisher. Elle a été introduite pour le cas de deux classes par Fisher et généralisée par Rao au cas de plusieurs classes. Une matrice rectangulaire W représentant la transformation linéaire de l'espace des primitives de dimension D vers un espace réduit de dimension d $(d << D)$ est obtenue par la maximisation du critère de Fisher qui est défini comme le rapport de la trace de la matrice inter classe par celle de la matrice intra classe. L'estimation de ce critère peut être obtenue à partir d'un échantillon de vecteurs dimension D , où k représente le label des vecteurs qui sont de dimension D. Les estimateurs respectifs de la matrice inter classe S_b at de la matrice intra classe S_w sont souvent prises comme suit en fonction d'un échantillon d'apprentissage $\{X_i^k \in \Re^D, 1 \leq i \leq N, 1 \leq k \leq K\}$ supervisé que nous notons par:

$$\hat{S}_b = \sum_{k=1}^{K} \pi_k (\mu_k - \mu)(\mu_k - \mu)^t \ et\ \hat{S}_w = \sum_{k=1}^{K} \pi_k \frac{1}{N_k} \sum_{i=1}^{N_k} (X_i^k - \mu_k)(X_i^k - \mu_k)^t \tag{1}$$

avec

$$\mu = \sum_{k=1}^{K} \frac{1}{N_k} \sum_{i=1}^{N_k} X_i^k \ et\ \mu_k = \frac{1}{N_k} \sum_{i=1}^{N_k} X_i^k \tag{2}$$

représentent respectivement la moyenne empirique globales et la moyenne empirique de chaque classe. La matrice W de réduction de dimension de Fisher est obtenue par les d vecteurs propres correspondant aux d plus grandes valeurs propres de la matrice $(\hat{S}_b)(\hat{S}_w)^{-1}$. Il est à noter que les deux matrices de dispersion sont définies seulement à partir des moments d'ordre inférieur ou égal à 2. Par conséquent, dans les situations un peu complexe telles que le cas multimodale d'une même classe, les moments d'ordre faible ne peuvent pas décrire complètement sa dispersion statistique. Dans ce contexte l'analyse discriminante basée sur le critère de Fisher n'est plus capable de donner satisfaction. Dans l'objectif de s'affranchir de cette limitation, des distance entre les densités de probabilité conditionnelles pondérées par les probabilités a priori ont été suggérées dans la littérature. Dans ce qui suit, nous nous intéressons à la Distance de Patrick Fisher qui s'écrit comme suit :

$$d_2(\pi_1 f_1, \pi_2 f_2) = (\int_{\Re^D} |\pi_1 f_1 - \pi_2 f_2|^2 dx)^{1/2} \tag{3}$$

Alain Hillion et al ont proposé d'étudier dans le cadre de la reconnaissance de formes de textures (images de la mer). Les performances d'un tel extracteur linéaire par l'évaluation des probabilités d'erreur a posteriori. L'estimateur de cette distance introduit par Patrick et Fisher, est obtenu en remplaçant dans l'expression de la distance les densités de probabilité conditionnelles par leur estimateur de type méthode du noyau. En considérant le noyau gaussien, l'expression de la distance devient une somme finie de fonctions de Gauss.

3. Un nouvel estimateur de la distance de Patrick Fisher

L'extension de la méthode décrite à une d-réduction multidimensionnelle semble difficile. Nous proposons ici l'extension de cette méthode à la réduction vectorielle par l'utilisation de l'estimateur des densités de probabilité par les fonctions orthogonales. Cet estimateur suppose que la densité de probabilité appartient à un espace de Hilbert muni d'une base de fonctions orthogonales. L'estimateur s'écrit :

$$\hat{f}(x) = \sum_{\ell=1}^{m_N} \hat{a}_{N,\ell}\, e_\ell(x) \tag{4}$$

Où $\hat{a}_{N,\ell}$ sont des estimateurs des coefficients de Fourier s'exprimant de la manière suivante:

$$\hat{a}_{N,\ell} = \frac{1}{N}\sum_{j=1}^{N} e_\ell(X_j),\ \{e_\ell(x)\}_\ell \ est\ une\ base\ de\ l'espace\ de\ Hilbert\ L^2[X] \tag{5}$$

m_N est le paramètre de troncature de la série de Fourier. Ce paramètre joue le rôle d'un facteur de lissage similaire à la variance du noyau dans la méthode de Parzen appelé souvent pas de l'estimateur. La convergence en moyenne quadratique intégrée de l'estimateur des fonctions orthogonales, est obtenue quand la suite m_N est équivalente à $N^{-1/s}$ avec $s > 2$. Dans ce qui suit nous prendrons cette inégalité comme règle même si nous nous trouvons dans un contexte de classification (plusieurs classes). Il est bien connu qu'il est aussi possible d'exprimer l'estimateur par les fonctions orthogonales comme un cas plus général que celui des densités de probabilité à noyau :

$$\hat{f}(x) = \frac{1}{N}\sum_{j=1}^{N} K_{m_N}(x, X_j) \tag{6}$$

Ainsi, le noyau généralisé $K_{m_N}(x, X_j)$ pour une base de fonctions orthogonales , s'écrit de la manière suivante:

$$K_{m_N}(x, X_j) = \sum_{\ell=1}^{m_N} e_\ell(x) e_\ell(X_j) \tag{7}$$

Proposition 1: Considérons la suite d'entiers m_k, le noyau K défini à partir de la base de fonctions $e_i(x)$ vérifie ce qui suit :

$$K_{min(m_1,m_2)}(y, z) = \int_X K^*_{m_1}(x, y) K_{m_2}(x, z)dx \tag{8}$$

Proposition 2: La quantité suivante est un estimateur sans biais de la distance de Patrick Fisher

$$\hat{d}_p = \frac{1}{N^2}\{\sum_{l=1}^{N_1}\sum_{j=1}^{N_1} K_{m_1}(X_l^1, X_j^1) + \sum_{l=1}^{N_2}\sum_{j=1}^{N_2} K_{m_2}(X_l^2, X_j^2) - 2Re[\sum_{l=1}^{N_1}\sum_{j=1}^{N_2} K_{min(m_1,m_2)}(X_l^1, X_j^2)]\} \tag{9}$$

Cet estimateur de distance donne des réductions de dimension dans le cas de deux classes. Le cas du multi classes peut être traité de manière analogue en maximisant dans l'espace réduit la mesure de dépendance probabiliste qui s'écrit comme suit :

$$I_2 = \sum_{k=1}^{K} \pi_k (\int_X |f_k(x) - f(x)|^2 dx)^{1/2} \tag{10}$$

Les fonctions densités de probabilité conditionnelles à chaque classe sont remplacées dans l'expression de I_2 par sa série de Fourier tronquée.

4. Les simulations et les résultats expérimentaux

La meilleure performance de la distance de Patrick Fisher en analyse discriminante relativement à celle obtenue par les matrices de dispersion peut être montrée par simulation grâce à l'estimateur proposée dans cet article. Pour cela, nous générons N_1 observations résultant d'un vecteur aléatoire gaussien de dimension 3 et N_2 observations d'un vecteur aléatoire non gaussien de même dimension de support $D = \{u \in \Re^3 \setminus \|u\| > 2\}$. Ces observations sont représentées dans la figure 1.

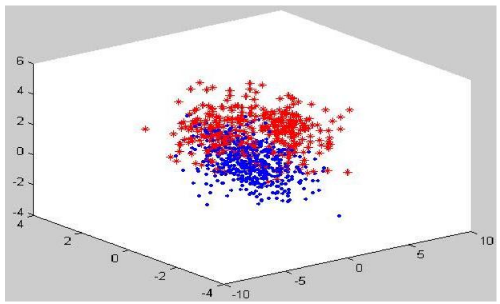

Figure 1. Une simulation d'un mélange non gaussian

La figure 2.a présente le résultat d'une réduction de dimension de paramètre $D = 3$ et $d = 2$ obtenu à la suite de l'application de la méthode de Fisher. Le résultat du même type de réduction de dimension obtenu par l'estimateur proposé de la distance de Patrick Fisher est présenté dans la figure 2.b.

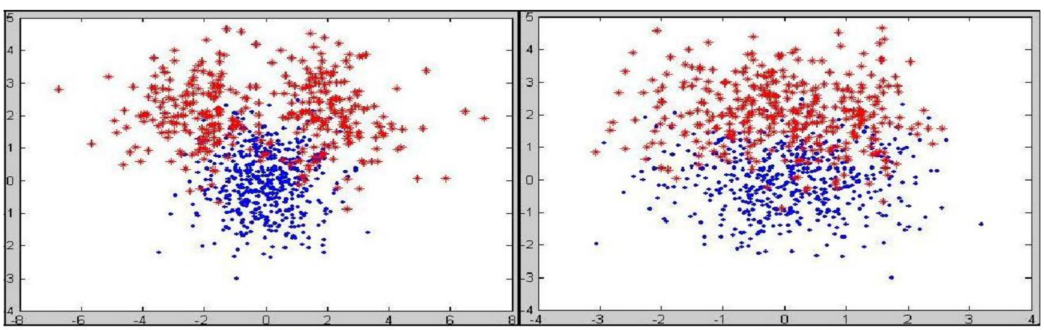

Figure 2.a **Figure 2.b**

Figure 2. Projection des données simulées sur l'espace réduit

2.a) suivant la méthode de *Fisher* **2.b)** suivant la méthode de *Patrick Fisher*

Il ressrot de ces illustrations que la distance de Patrick Fisher fournit un plan discriminant meilleur que celui de Fisher au sens d'une meilleure séparabilité des classes. La reconnaissance de visage est actuellement l'un des sujet d'applications les plus actifs en reconnaissance de formes: qui est utile pour la biomètrie. Le nombre de descripteurs recquis est souvent très élevé. Pour réussir une telle tâche, une réduction de dimension est nécessaire pour la convergence des estimateurs. L'application de la distance de Patrick Fisher trouve son intérêt dans ce type d'opération. La classification est ensuite opérée sur l'espace réduit.

Figure 3. un échantillon de visage en différant phase de pré-traitement

Dans la figure 3 nous représentons quelques prétraitements opérés sur un visage extrait de la base de données afin de préparer la phase d'extraction de primitives.

5. Conclusion

Dans ce travail, nous avons introduit deux nouveaux estimateurs. Un premier estime la distance de Patrick Fisher et un second fournit un estimateur pour la mesure de dépendance probabiliste. Ils se basent sur la méthode d'estimation des densités de probabilité par les fonctions orthogonales. Nous avons commencé par rappeler la limitation des méthodes basées sur les matrices de dispersion puis nous avons fait remarquer que l'extension de l'extracteur linéaire scalaire obtenu par l'estimateur à noyau de la distance de Patrick Fisher peut être étendue naturellement au cas multidimensionnel lorsque l'on considére le type du nouvel estimateur à base de fonctions orthogonales. Leur application prometteuse en reconnaissance de formes est illustrée sur le problème difficil de la reconnaissance faciale. En perspective de ce travail une étude sur l'analyse discrimante tenant compte d'information sur le support des données à classifier par ces deux estimateurs sera menée. En effet, pour chaque type d'espace de Hilbert des bases de fonctions adaptées (Legendre Laguerre, trigonométriques, Hermites..) seront introduites. Des études asymptotiques par simulations feront l'objet des travaux futures pour montré la consistance des deux estimateurs.

Bibliographie

[1] Marco Loog and al, Multiclass Linear Dimension Reduction by Weighted Pairwise Fisher Criteria, IEEE transaction on PAMI Vol 23 N 7 Juilly 2001

[2] E.A. Patrick and F.P. Fisher II, "Nonparametric feature selection", IEEE Trans. On Inf. Theory, vol. IT-15, pp.577-584, 1969.

[3] A. Hillion and al. A non parametric approach to linear feature extraction; Application to classification of binary synthetic textures", 9th International Conference on Pattern Recognition, ICPR pp.1036-39, 1988.

[4] B.W Silverman, Density Estimation for Statistics and Data Analysis, London, Charman and Hall, 1986.