

Sélection de variables par SVM : application à la détection de piétons

SVM Variable selection with application to pedestrian detection

Alain Rakotomamonjy

Frédéric Suard

Laboratoire Perception, Systèmes et Informations, FRE CNRS 2645

INSA de Rouen

Avenue de l'Université 76801 Saint Etienne du Rouvray

alain.rakotomamonjy@insa-rouen.fr

Résumé

La prolifération de systèmes d'acquisition et de stockage des données permet d'augmenter la quantité de données disponibles pour la résolution d'un problème de discrimination à partir d'exemples. Par ailleurs, chaque exemple peut être décrit par un grand nombre de caractéristiques. La sélection de variables vise à extraire parmi ces caractéristiques celles qui sont les plus pertinentes vis-à-vis du problème à traiter. Dans cet article, nous proposons un critère d'évaluation d'une variable pour les Séparateurs à Vaste Marge. Celui-ci est basé sur la sensibilité de la marge à une variable. Pour les SVM multiclassées, une extension naturelle de ce critère est également proposée. Après avoir décrit les détails techniques de l'obtention du critère, l'efficacité de la méthode de sélection de variables a été testée de manière convaincante sur des problèmes-jouets. La reconnaissance de piétons à partir d'images est typiquement le genre d'application pour laquelle la sélection de variables peut être bénéfique car elle permet tout au moins d'aboutir aux mêmes performances de reconnaissance à partir d'un nombre de caractéristiques moindres. Ainsi, une étude sur l'apport de la sélection de variables a été menée et démontre effectivement l'intérêt de choisir les caractéristiques de manière pertinente.

Mots Clef

Apprentissage, Séparateurs à Vaste Marge, Sélection de Variables, Piétons.

Abstract

Nowadays, to solve a learning from examples problem, a large amount of data can be available due to the proliferation of acquisition systems. Besides, more and more features characterize each example. Variable selection algorithms address the problem of selecting the input variables that are most predictive for a given task. In this paper, we present a variable evaluation criterion for Support Vector Machines which is based on the influence of a given variable on the margin. Technical details concerning the way

to derive this criterion are then given. Efficiency of the proposed method has been tested on some toy problems which show that in a classification task, our algorithm is able to retrieve the most predictive variables among a set of noisy variables. Our variable selection method has then been applied to pedestrian detection in images. Such a problem is typically a problem that can benefit of variable selection and our experiments demonstrate that a SVM based variable selection make it possible to reduce the number of variables without compromising the recognition rate. This is of importance for such a real-time application.

Keywords

Learning, Support Vector Machines, Variable Selection, Pedestrian Detection

1 Introduction

L'objectif d'un problème de discrimination à partir d'un ensemble d'exemples est d'inférer un lien entre les caractéristiques de chaque exemple et leur catégorie usuellement définie par un label. Actuellement, de par la facilité d'acquisition et de stockage des données, les problèmes réels de reconnaissance de formes ou de discrimination sont de plus en plus complexes et font intervenir de plus en plus de variables caractérisant un exemple. Si intuitivement il semble naturel que l'augmentation du nombre caractéristiques ne devrait pas nuire à la qualité de la discrimination, en pratique cela s'avère être un problème majeur. En effet, le succès d'un problème de reconnaissance de formes est fortement lié à la qualité des données et des variables qui les caractérisent. La présence de variables redondantes, bruitées et peu corrélées avec le label d'un exemple rend le processus d'apprentissage plus difficile.

La sélection de variables est un problème consistant à identifier les variables pertinentes par rapport au problème considéré (dans notre cas la discrimination) [10, 7]. Si la sélection des variables est réalisée de manière appropriée, la performance du modèle de discrimination peut être fortement améliorée tout en réhaussant l'interprétabilité du mo-

dèle. Par ailleurs, le fait de réduire le nombre de variables permet également de réduire le temps de calcul du label d'un nouvel exemple présenté au modèle d'inférence. Typiquement un problème de sélection de variables nécessite la définition de plusieurs critères :

- la fonction d'évaluation : cette fonction doit permettre de mesurer la pertinence d'un sous-ensemble de variables par rapport au problème considéré. Plusieurs types de critères ont été proposés dans la littérature et sont par exemple basés sur des méthodes telles que l'information mutuelle entre variables [1] ou encore la sensibilité de la sortie d'un réseau de neurones à une variable donnée [11].
- la méthode de recherche : celle-ci définit la façon permettant d'explorer l'espace des sous-ensembles de variables. Cet espace étant de dimension 2^d (d étant le nombre de variables), un parcours exhaustif de l'ensemble n'est possible que lorsque le nombre de variables est faible (de l'ordre de 10). Ainsi, les techniques les plus fréquemment utilisées sont des algorithmes sous-optimaux tels que les algorithmes génétiques [17, 21] ou des algorithmes itératifs [4]
- le critère d'arrêt : étant donné que l'exploration exhaustive de l'espace des variables est prohibitive en temps de calcul, il est nécessaire de définir un critère d'arrêt de l'exploration. Ainsi, après arrêt de l'exploration, le "meilleur" sous-ensemble de variables est celui qui optimise la fonction d'évaluation.

Dans cet article, notre contribution réside principalement dans la définition d'un nouveau critère d'évaluation d'une variable par rapport à sa pertinence vis-à-vis d'un problème de discrimination par Séparateur à Vaste Marge (SVM) [16, 18]. Bien que théoriquement considéré comme insensible à la dimensionnalité du problème de discrimination, il s'avère en pratique que les SVM souffrent également de la redondance et de la non-pertinence de certaines variables. Ainsi, quelques algorithmes de sélection de variables ont déjà été proposés. Celui de Weston et al. utilise comme critère d'évaluation une borne sur l'erreur de rééchantillonnage et propose d'explorer l'espace des variables par une méthode de descente de gradient [19]. La méthode décrite par Guyon et al. consiste à éliminer itérativement les variables qui font le moins décroître la marge [8]. Dans notre algorithme, la pertinence d'une variable est estimée à travers son influence sur la marge.

Le cadre applicatif qui nous intéresse est la détection de piétons dans une image. Pour la conception de système d'aide à la conduite, ce problème est de première importance afin d'éviter des situations dangereuses. Par ailleurs, une contrainte de type temps réel est nécessaire dans cette application. Ainsi, une fois que les caractéristiques d'un piéton sont extraites d'une image, un processus de sélection de variables peut être bénéfique en deux points : identifier les variables pertinentes permet d'augmenter la performance du détecteur et de réduire le temps de calcul. Dans cet article, nous reprendrons l'étude menée au MIT sur la

détection de piétons [14] en y appliquant notre algorithme de sélection de variables.

Cet article est structuré comme suit : tout d'abord, nous présentons un bref descriptif des Séparateurs à Vaste Marge. Ensuite, nous détaillons notre algorithme de sélection de variable ainsi que le critère d'évaluation proposé. Une extension aux cas multiclassés est également proposée. Après avoir testé brièvement la pertinence de notre algorithme sur des problèmes jouets, nous appliquons l'algorithme au problème de détection de piétons. Enfin, la dernière partie de l'article résume notre travail, présente une discussion ainsi que les futurs travaux.

2 Les Séparateurs à Vaste Marge

Cette section décrit brièvement le principe des Séparateurs à Vaste Marge (SVM).

2.1 Cas linéairement séparable

Étant donné un ensemble de points d'apprentissage $\{x_i, y_i\}_{i=1}^{\ell}$ où chaque x_i appartient à \mathbb{R}^d et y_i à $\{-1, 1\}$, y_i définissant la classe d'un exemple donné, l'objectif des SVM est de trouver un hyperplan permettant de séparer l'ensemble d'apprentissage de sorte que tous les points d'une même classe soient d'un même côté de l'hyperplan. Cela équivaut (après un éventuel normalisation) à trouver un hyperplan $f(x)$ défini par $w \in \mathbb{R}^d$ et $b \in \mathbb{R}$ tel que :

$$y_i(\langle w, x_i \rangle + b) \geq 1 \quad i = 1 \dots \ell$$

Parmi l'ensemble des hyperplans satisfaisant ces conditions, les SVM cherchent celui qui maximise la distance entre l'hyperplan et les points les plus proches de chaque classe, cette distance étant définie comme $2/\|w\|$. Ainsi le problème global équivaut à :

$$\begin{aligned} \min_{w, b} & \frac{1}{2} \|w\|^2 \\ \text{tq} & y_i f(x_i) \geq 1 \quad i = 1 \dots \ell \end{aligned} \quad (1)$$

Dans la mesure où $\|w\|^2$ est un critère convexe et l'ensemble des contraintes de ce problème sont linéaires, ce problème peut se résoudre en utilisant les multiplicateurs de Lagrange α et le dual lagrangien de ce problème. Cela permet d'écrire que le vecteur w est :

$$w = \sum_{i=1 \dots \ell} \alpha_i^* y_i x_i$$

où les α_i^* sont la solution du problème suivant :

$$\begin{aligned} \min_{\alpha} & \frac{1}{2} \sum_{i, j} \alpha_i \alpha_j y_i y_j \langle x_i, x_j \rangle - \sum_{i=1}^{\ell} \alpha_i \\ \text{tq} & \sum_{i=1}^{\ell} \alpha_i y_i = 0 \\ & \alpha_i \geq 0 \quad \forall i \end{aligned} \quad (2)$$

Ce dernier est un problème d'optimisation quadratique standard qui se résout maintenant grâce à des outils courants. Finalement l'hyperplan de décision est :

$$f(x) = \sum_{i=1}^{\ell} \alpha_i y_i \langle x_i, x \rangle + b \quad (3)$$

où b est le multiplicateur de Lagrange associé à la contrainte $\sum_{i=1}^{\ell} \alpha_i y_i = 0$ du problème dual. La classe d'un nouveau point x est ensuite donné par le signe de $f(x)$.

2.2 Cas non-linéaires

Le paragraphe précédent décrit le principe des SVM dans le cas où les données sont linéairement séparables. Cependant, dans la plupart des problèmes réels, ce n'est pas toujours le cas et il est donc nécessaire de contourner ce problème. Dans ce but, l'idée est de projeter les points d'apprentissage dans un espace \mathcal{H} donné grâce à une fonction $\Phi(x)$ et d'appliquer la même méthode d'optimisation de la marge dans l'espace \mathcal{H} . Ce dernier problème s'écrit simplement puisque d'après les équations (2) et (3) les points x_i n'interviennent dans la solution qu'à travers leurs produits scalaires. Ainsi le cas non-linéaire s'obtient aisément en remplaçant $\langle x_i, x_j \rangle$ par $k(x_i, x_j) = \langle \Phi(x_i), \Phi(x_j) \rangle$. Dans ce cas, la frontière de décision devient :

$$f(x) = \sum_{i=1}^{\ell} \alpha_i y_i k(x_i, x) + b$$

Comme on peut le constater, la projection des point x_i dans \mathcal{H} se fait de manière implicite à travers la définition de la fonction $k(x, y)$. Si cette fonction $k(x, y)$ satisfait aux conditions de Mercer [12, 16] alors il existe une fonction Φ telle que $k(x, y) = \langle \Phi(x), \Phi(y) \rangle$. L'ensemble des fonctions noyaux $k(x, y)$ admissibles pour les SVM sont donc celles-ci. En pratique, les noyaux les plus utilisés sont les noyaux gaussiens ou les noyaux polynomiaux définis par :

- noyau gaussien de largeur de bande σ :

$$k(x, y) = e^{-\frac{\|x-y\|^2}{2\sigma^2}}$$

- noyau polynomial d'ordre p :

$$k(x, y) = (\langle x, y \rangle + 1)^p$$

2.3 Cas non séparables

Dans le cas où l'ensemble d'apprentissage n'est pas séparable, le domaine du problème d'optimisation comme défini à l'équation (1) est vide et il n'admet donc pas de solution. Ainsi, pour traiter ce cas, il est nécessaire d'introduire des variables de relâchement au niveau des contraintes, et le problème devient :

$$\min_{w, b, \xi_i} \frac{1}{2} \|w\|^2 + C \sum_{i=1}^{\ell} \xi_i^2 \quad \text{tq} \quad y_i f(x_i) \geq 1 - \xi_i \quad i = 1 \dots \ell \quad (4)$$

où ξ_i est la variable de relâchement d'une contrainte et C un coefficient de pénalisation du relâchement. Le traitement de ce problème se fait de manière simple en montrant que l'ajout de la pénalisation quadratique des points mal classés équivaut à traiter le cas séparable en remplaçant $k(x_i, x_j)$ par [5]:

$$k(x_i, x_j) + \frac{1}{C} \delta_{i,j}$$

où $\delta_{i,j}$ est le symbole de Kronecker. Notons que la pénalisation quadratique peut être remplacée par une pénalisation linéaire.

2.4 SVM multiclass

A l'origine, les SVM ont été conçus essentiellement pour les problèmes à 2 classes, cependant plusieurs approches permettant d'étendre cet algorithme aux cas à N classes ont été proposées [9] :

- l'approche la plus naturelle est d'utiliser cette méthode de discrimination binaire et d'apprendre N fonctions de décision $\{f_m\}_{m=1 \dots N}$ permettant de faire la discrimination entre chaque classe de toutes les autres. L'affectation d'un nouveau point x à une classe C_i se fait par la relation [2] :

$$i = \arg \max_{m=1 \dots N} f_m(x)$$

- la deuxième méthode, celle que nous utilisons dans cette communication, est une méthode dite de *un contre un*. Au lieu d'apprendre N fonctions de décisions, ici chaque classe est discriminée d'une autre. Ainsi, $\frac{N(N-1)}{2}$ fonctions de décisions sont apprises et chacune d'entre elles effectue un vote pour l'affectation d'un nouveau point x . La classe de ce point x devient ensuite la classe majoritaire après le vote [6].
- la dernière méthode est une approche étendant la notion de marge aux cas multiclass [13, 20]. Cette formulation intéressante permet de poser un problème d'optimisation unique justifié par des considérations théoriques. Le problème fait intervenir N fonctions de décision et il s'écrit dans le cas où les données sont non séparables :

$$\min_{w, b, \xi_i^m} \sum_{o < m}^N \|w_m - w_o\|^2 + C \sum_{i=1, m \neq y_i}^{\ell} \xi_i^m \quad \text{tq} \quad \begin{aligned} f_{y_i}(x_i) &\geq f_m(x_i) + 2 - \xi_i^m \\ \sum_{m=1}^N w_m &= 0 \\ \xi_i^m &\geq 0 \quad i = 1 \dots \ell, m = 1 \dots N \end{aligned}$$

Cet algorithme est très gourmand en temps de calcul et en espace mémoire ce qui fait qu'il reste peu utilisé dans les cas réels.

3 Selection de Variables par SVM

Ce paragraphe décrit plus en détails notre contribution à l'état de l'art de la sélection de variable, à travers la définition d'un nouveau critère d'évaluation basé sur les SVM.

3.1 les techniques de sélection de variables

Dans un problème de discrimination, le but de la sélection de variables est d'éliminer les variables redondantes ou peu informatives. Dans cet article, nous nous focalisons principalement sur le problème de classement de variables. Le

critère original que nous introduisons est basé sur la sensibilité de la marge à une variable donnée. L'heuristique se sachant derrière ce critère peut être résumée de la façon suivante :

Une variable peu informative et donc peu importante pour la fonction de décision, est une variable auquelle la marge $2/\|w\|$ est peu sensible.

Nous détaillerons ce critère dans le paragraphe suivant. Pour le parcours de l'espace des sous-ensemble de variables, nous nous basons sur une méthode déjà largement éprouvée consistant à effectuer une suppression séquentielle des variables. L'algorithme de cette méthode est :

1. Initialisation **trie** = []; **var** = $[1 \dots d]$
2. Répéter
 - (a) évaluer la marge $2/\|w\|$ de la fonction de décision utilisant les variables **var**
 - (b) pour toutes les variables dans **var**, calculer la sensibilité de la marge par rapport à chaque variable
 - (c) sélectionner la variable **b** qui minimise la sensibilité
 - (d) classer cette variable : **trie** = [**b** **trie**];
 - (e) supprimer **b** de la liste **var**
3. jusqu'à ce que toutes les variables soient triées

où **var** et **trie** sont les vecteurs de variables respectivement à trier et déjà trié (par ordre décroissant de pertinence).

3.2 Sensibilité de la marge par rapport à une variable

Pour mesurer la sensibilité de la marge par rapport à une variable, nous introduisons un facteur d'échelle ν et allons calculer la norme du gradient de la marge par rapport à ce facteur d'échelle. Ce vecteur ν est en fait un vecteur unitaire pondérant chaque variable et de ce fait, le noyau $k(x,y)$ dans l'équation (2) et (3) devient :

$$k(\nu \cdot x, \nu \cdot y)$$

où l'opérateur \cdot est le produit vectoriel terme à terme.

Ainsi, pour évaluer la sensibilité de la marge en fonction d'une variable v , on s'intéresse à la quantité :

$$\left| \frac{\partial M}{\partial \nu_v} \right| = \left| \frac{\partial}{\partial \nu_v} \frac{2}{\|w\|} \right| = \left| -\frac{2}{\|w\|^2} \frac{\partial \|w\|}{\partial \nu_v} \right|$$

D'après l'algorithme, la suppression d'une variable est décidée par comparaison aux autres variables, toutes les valeurs constantes en fonction de v dans $\frac{\partial M}{\partial \nu_v}$ peuvent être

abandonnées. Ainsi, le critère d'évaluation que nous étudions devient :

$$\left| \frac{\partial \|w\|^2}{\partial \nu_v} \right| = \left| \frac{\partial}{\partial \nu_v} \sum_{i,j} \alpha_i^* \alpha_j^* y_i y_j k(\nu \cdot x_i, \nu \cdot x_j) \right|$$

La dérivation de cette expression peut sembler problématique car il paraît intuitif que les α_i^* dépendent de ν à travers le problème d'optimisation (2). La proposition suivante de Chapelle et al. permet de calculer simplement cette dérivée.

Proposition : *Chapelle et al, 2002 [3]*

Soit un vecteur v_θ et une matrice carré P_θ dépendant de manière continu d'un vecteur de paramètre θ . Considérons la fonction $F(\theta)$

$$F(\theta) = \min_{x \in G} \frac{1}{2} x^T P_\theta x - x^T v_\theta \quad (5)$$

où $G = \{x : b^T x = c, x \geq 0\}$ où b est un vecteur. Si \bar{x} est le vecteur x pour lequel le minimum de $F(\theta)$ est atteint et \bar{x} est unique alors :

$$\frac{\partial F(\theta)}{\partial \theta} = \frac{1}{2} \bar{x}^T \frac{\partial P_\theta}{\partial \theta} - \bar{x}^T \frac{\partial v_\theta}{\partial \theta} \bar{x}$$

■

En utilisant cette proposition conjointement au fait que les valeurs objectifs des problèmes d'optimisation (1) et (2) sont opposés (les deux problèmes sont duaux au signe près et convexes), on peut montrer que :

$$\left| \frac{\partial \|w\|^2}{\partial \nu_v} \right| = \left| -2 \sum_{i,j} \alpha_i^* \alpha_j^* y_i y_j \frac{\partial k(\nu \cdot x_i, \nu \cdot x_j)}{\partial \nu_v} \right|$$

Ainsi dans le cas de l'utilisation d'un noyau gaussien, le critère d'évaluation d'une variable v est donc :

$$\left| \frac{2}{\sigma^2} \sum_{i,j} \alpha_i^* \alpha_j^* y_i y_j \nu_v \left(x_i^{(v)} - x_j^{(v)} \right)^2 k(\nu \cdot x_i, \nu \cdot x_j) \right|$$

en rappelant que ν est un vecteur unitaire, ce critère se simplifie en :

$$\left| \frac{2}{\sigma^2} \sum_{i,j} \alpha_i^* \alpha_j^* y_i y_j \left(x_i^{(v)} - x_j^{(v)} \right)^2 k(x_i, x_j) \right|$$

3.3 Extension aux cas multiclasse

Si le problème multiclasse est traité à partir de multiples classifieurs binaires comme dans le cas des SVM *un contre tous* et *un contre un* alors la notion de marge globale n'existe pas. Cependant, il est naturel de considérer comme critère d'évaluation d'une variable, son influence sur l'ensemble

des marges des différents problèmes de discrimination binaires. Ainsi, pour le cas *un contre tous* et *un contre un*, le critère d'évaluation est :

$$\left| \frac{\partial \sum_u M_u}{\partial \nu_v} \right|$$

où M_u est la marge du problème binaire u (dans le *un contre tous*, on a k classifieurs binaires tandis que dans le cas *un contre un* $k(k-1)/2$ classifieurs doivent être considérés. Comme dans le cas binaire, il est possible de montrer que ce critère équivaut à :

$$\left| \sum_u \sum_{i,j} \frac{2}{\sigma^2} \alpha_{i,u}^* \alpha_{j,u}^* y_i y_j (x_i^{(v)} - x_j^{(v)})^2 k(x_i, x_j) \right|$$

où les $\alpha_{i,u}$ correspondent aux multiplicateurs de Lagrange associés à un problème binaire u donné et un exemple i .

4 Expériences et applications

Dans ce paragraphe, sont présentés les résultats des expériences de sélection de variables que nous avons mené.

4.1 Problèmes-jouet

Deux problèmes de sélections de variables à partir de données synthétiques ont été traités ici. Le premier est un problème de classification à deux classes non linéaire tandis que le deuxième est un problème multiclassé. Dans chaque cas, les exemples sont caractérisés par 52 variables dont seules deux d'entre elles sont pertinentes pour la discrimination. L'objectif de l'algorithme de sélection de variables est donc de classer ces deux variables comme étant les variables auxquels la marge est plus sensible.

Nous avons évalué la capacité de l'algorithme proposé à sélectionner les variables pertinentes dans différentes situations. Pour cela, le critère que nous avons mesuré est l'erreur de généralisation d'un classifieur SVM obtenu à partir des exemples caractérisés seulement par les deux variables retenues comme étant les plus pertinentes par l'algorithme. Cette mesure de performance est quelque peu biaisée à la hausse dans la mesure où on suppose que le classifieur connaît le nombre optimal de variables à retenir pour l'apprentissage. La méthodologie utilisée est la suivante : 10000 et 3000 exemples ont respectivement été générés pour le problème à 2 classes et multiclassés. Seulement une faible partie d'entre eux tirés aléatoirement, font partie de l'ensemble d'apprentissage tandis que le reste des données forme l'ensemble de test. Les données d'apprentissage sont ensuite normalisées de façon à obtenir une moyenne nulle et une variance unitaire. De même, les points de test sont ajustés grâce à ces mêmes paramètres de normalisation. Une fois les données prétraitées, l'algorithme de sélection de variables a été lancé. Que ce soit pour la sélection de variables ou pour la discrimination SVM, un noyau gaussien de largeur de bande $\sigma = 3$ et $\sigma = 1$ a été utilisé respectivement pour le problème à 2 classes et multiclassé.

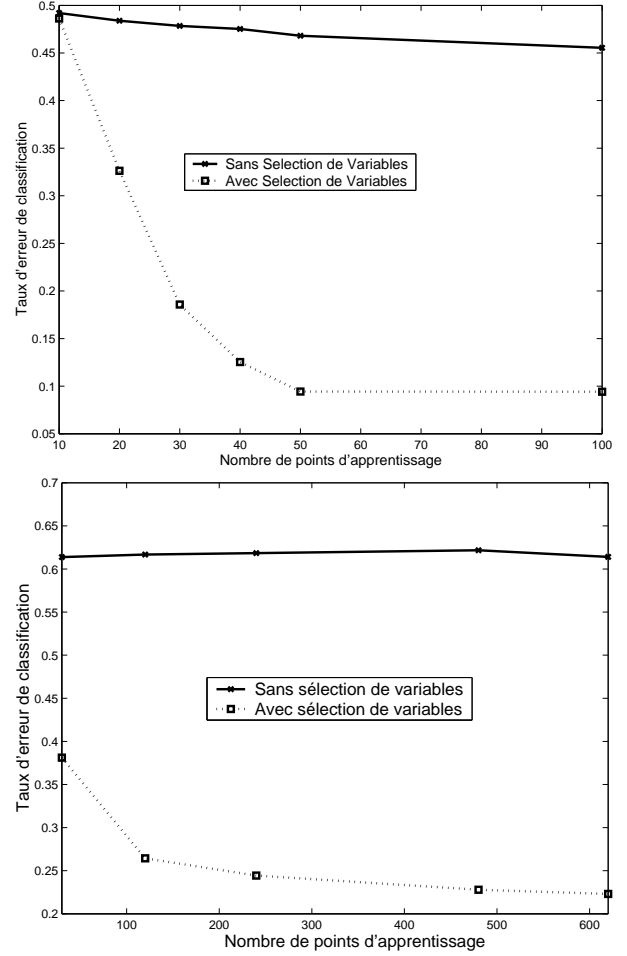


FIG. 1 – Performance des SVM pour des problèmes de classification. (haut) Classification binaire. (Bas) Classification multiclassé. Dans chaque cas, les deux courbes représentent le taux d'erreur de classification sans sélection de variables (en trait plein) et en associant au classifieur SVM l'algorithme de sélection de variables basé sur la sensibilité de la marge (trait pointillé).

Le paramètre de pénalisation C a été fixé à une valeur suffisamment élevée pour que l'erreur d'apprentissage reste faible (respectivement $C = 1000$ et $C = 100$). Cent expériences dans le cas à 2 classes ont été effectuées tandis que pour le problème multiclassé, nous avons restreint ces expériences à 10 essais à cause de la forte complexité temporelle de l'algorithme. La figure (1) représente la moyenne de l'erreur en généralisation (mesurée sur les données de test) en utilisant l'algorithme de sélection de variables conjointement à un classifieur SVM et un classifieur SVM sans sélection de variables pour les 2 problèmes de classification. Ces courbes montrent clairement que lorsqu'ils sont confrontés à un grand nombre de variables peu pertinentes les SVM ont tendance à sur-apprendre et leur capacité de généralisation s'en trouve affectée. Le fait de rajouter un algorithme de sélection de variables permet de faire face à ce problème. Au fur à mesure que le



FIG. 2 – Exemples de piétons de la base de données.

nombre d'exemples d'apprentissage croît, l'algorithme de sélection de variable est de plus en plus performant et semble être capable de classer les 2 variables informatifs comme étant les plus pertinentes. Ces problèmes-jouet confirment donc que la sensibilité de la marge à une variable est un "bon" critère d'évaluation de la pertinence d'une variable. Le lecteur peut arguer que ces expériences sont justes des expériences-jouets et ne doivent être considérer que en tant que telles. Cependant, une étude plus approfondie de cette méthode de sélection de variables peut être trouvée en [15]. Cette dernière référence décrit également des discussions concernant l'importance des paramètres C et σ du noyau gaussien.

4.2 Détection de piétons

La mise au point d'un système de détection de piétons robuste est un point de première importance pour les systèmes d'aide à la conduite. Actuellement, ce problème est loin d'être résolu et plusieurs axes de recherche sont examinés par différents laboratoires. Nous nous intéressons ici à une approche de détection basé sur l'apprentissage et non sur un modèle. Cette idée initialement développée par Papageorgiou et al. consiste à apprendre à un système la "forme" d'un piéton [14]. Le principe de la méthode consiste donc à créer une base d'exemples de piétons (suivant différentes poses) ainsi qu'une base d'images ne représentant pas des piétons. Dans cette expérience, la base de piétons que nous avons utilisée est celle mise à disposition par le MIT ainsi qu'un ensemble d'exemples acquis par notre soin. La taille de chaque image est de 128×64 et ces images sont de sorte que le piéton y soit centré.

De chaque image couleur, doivent être extraites des caractéristiques qui sont ensuite fournies à un algorithme d'apprentissage qui dans notre cas est un SVM. La représentation que nous avons choisie pour chaque image, est une décomposition en ondelettes sur une base redondante de Haar. Le lecteur désireux d'avoir plus de détails sur cette procédure peut se référer à l'article de Papageorgiou et al [14]. Ainsi, chaque canal de l'image RGB est décomposé pour obtenir une approximation de niveau 5. Les décompositions verticales, horizontales et diagonales de chaque

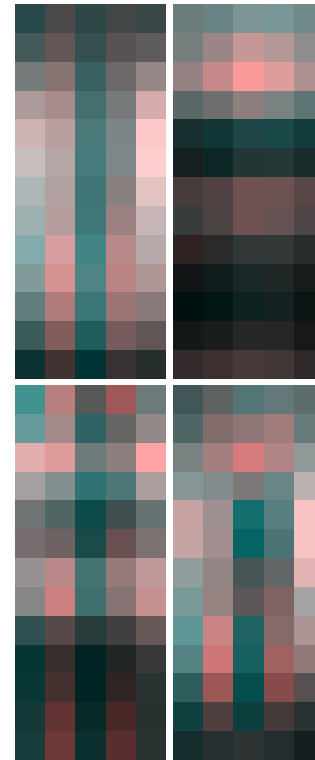


FIG. 3 – Coefficients en ondelettes moyens obtenus sur la base d'apprentissage. (haut-gauche) Verticale. (haut-droit) Horizontale. (bas-gauche) Diagonale. (bas-droit) Maximum des coefficients sur les 3 directions. Le niveau de gris indique l'amplitude du coefficient allant du plus faible pour le noir au plus élevé pour le blanc.

canal sont ensuite comparées et pour un coefficient donné, seul le coefficient le plus élevé en valeur absolue suivant les 3 couleurs est retenu. L'ensemble des coefficients est ensuite normalisé de sorte que la valeur maximale des coefficients soit 1. Cela faisant, chaque image est représentée par un vecteur de taille 195 suivant les 3 directions. La figure (3) représente l'amplitude moyen des coefficients d'ondelettes sur la base d'exemples. Malgré la faible résolution de l'image, on distingue nettement les contours des piétons notamment pour les coefficients verticaux et diagonaux. Notons que dans leur travail, Papageorgiou et al. développent un modèle de piétons à partir des coefficients en ondelettes qu'il appelle *wavelet pedestrian template*. Ils justifient l'utilisation de ce modèle par la nécessité de réduire la variabilité de la luminosité entre les images de piétons. A travers cela, leur idée est principalement de représenter une image de piétons par les variations de luminosité moyenne sur le support d'une ondelette. Ici, les coefficients de la décomposition sont utilisés tels quels. Cela permet de tenir compte lors de l'apprentissage de l'ensemble des niveaux de gris de l'image.

Notre base d'exemples est composée de 150 images de piétons et de 150 images de scènes routières. Pour l'application, elle est décomposée en deux parties de respectivement

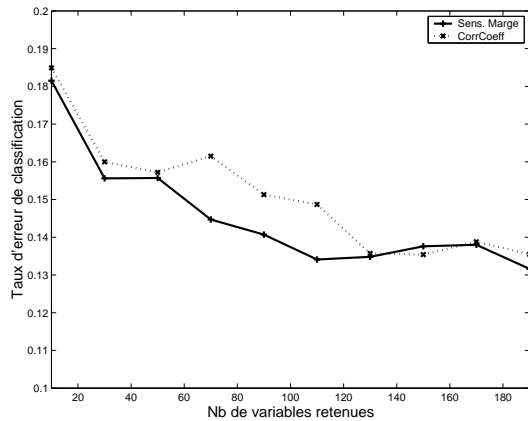


FIG. 4 – Taux d’erreur de classification en fonction du nombre de variables retenu : en trait pleins, notre critère d’évaluation, en pointillés, un critère d’évaluation basé sur le coefficient de corrélation.

100 et 50 images tirées aléatoirement formant ainsi la base d’apprentissage et la base de test. Le noyau utilisé pour la sélection de variables et pour la classification est un noyau polynomial d’ordre 2 tandis que le coefficient de pénalisation C est fixé à 1000. Le choix de ces paramètres ont été fixés de manière à obtenir un taux d’erreur en apprentissage faible.

L’expérience a été itérée cent fois en faisant varier le nombre de variables retenu par l’algorithme de sélection de variables. La figure (4) décrit le taux d’erreur moyen sur la base de test en fonction du nombre de variables. Les résultats obtenus par notre méthode sont également comparés à ceux d’un même algorithme de suppression séquentielle de variables pour lequel le critère d’évaluation est le coefficient de corrélation d’une variable avec la classe d’une image.

Cette figure montre que la performance de classification est maximale en utilisant l’ensemble des caractéristiques extraites de l’image. La stratégie utilisée pour la discrimination semble être intéressante puisque qu’elle permet d’atteindre un taux d’erreur de l’ordre de 13%. Cependant, au moins la moitié des variables peuvent être écartées sans compromettre la qualité de la discrimination. En effet, en ne retenant que 90 variables le taux d’erreur ne croît que de 1% tandis que l’utilisation de 50 variables affecte la performance (une différence de 2% apparaît). La comparaison entre les deux critères d’évaluation est en faveur de notre méthode originale lorsque le nombre de variables supprimés est de l’ordre de la moitié du nombre de variables total. Cela permet donc de conjecturer que les premières variables supprimées apporte peu à la discrimination.

Bien que la sélection de variables ne permet pas de réhausser la performance du classifieur, ce qui signifierait que l’ensemble des caractéristiques de l’image sont pertinentes pour la classification, la sélection de variables permet donc de réduire le nombre de variables nécessaires à la clas-

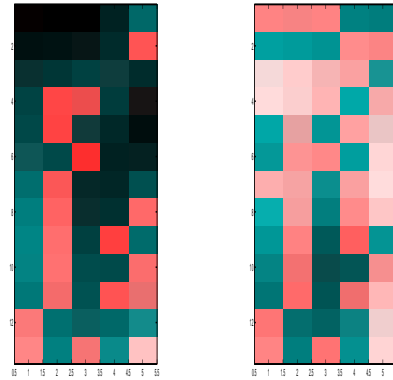


FIG. 5 – Exemples de variables sélectionnées par l’algorithme. (gauche) Variables sélectionnées, en blanc, suivant la décomposition verticale. (droite) Localisation de l’ensemble des variables sélectionnées suivant les 3 axes.

sification. Cela est de première importance pour ce type d’application avec une contrainte de temps réel. La réduction des caractéristiques discriminantes permet donc d’accélérer le processus de classification. La figure (5) donne un exemple de 50 variables sélectionnées par l’algorithme. Les variables issues de la décomposition verticale considérées comme pertinentes sont représentées en blanc. Nous avons également représenté les variables pertinentes en fonction de leur localisation sur l’image quelque soit la décomposition (verticale, horizontale ou diagonale).

En confrontant les résultats obtenus sur la figure (5) et les figures (2) et (3), on constate que les variables les plus pertinentes pour la détection de piétons semblent être celles qui caractérisent le contour du piéton (la colonne 2 et 4). Cela suggère donc une stratégie de discrimination des piétons que nous envisagerons dans un futur proche : une méthode basée sur l’image elle-même et les contours extraits de l’image.

5 Conclusions et perspectives

Nous avons présenté dans cet article une méthode originale de sélection de variables pour les Séparateurs à Vaste Marge. Cette méthode s’articule autour d’un critère d’évaluation basé sur la sensibilité de la marge à une variable et d’un algorithme de suppression séquentielle de variables. Bien que défini initialement pour les SVM à deux classes, le critère d’évaluation s’étend de manière naturelle aux SVM multiclassés basés sur les techniques de *un contre un* ou *un contre tous*.

L’algorithme de sélection de variables a ensuite été mis en oeuvre sur des problèmes-jouets où il a démontré toute son efficacité. En effet, au fur et à mesure de l’augmentation du nombre d’exemples dans la base d’apprentissage, les expériences montrent que l’algorithme est capable de trouver les variables pertinentes pour la discrimination.

L’application utilisée pour illustrer la méthodologie est la détection de piétons. Après avoir mis au point une stratégie de discrimination, nous avons été confrontés à un grand

nombre de variables discriminantes. Ainsi, nous avons pu tester l'efficacité de notre algorithme dans un cas réel. En effet, sans pour autant accroître les performances du classifieur, nous avons montré qu'il est possible d'obtenir des performances similaires en utilisant tout au plus la moitié des variables. Cela reste avantageux car la complexité temporelle de la détection s'en trouve réduite. La méthodologie que nous avons suivie pour cette application est celle décrite par Papageorgiou et al. dans leurs travaux sur les mêmes applications. Cependant, nous avons apporté une nouvelle contribution à cette approche dans la mesure où, dans nos travaux, les variables sont sélectionnées automatiquement à travers un algorithme et non manuellement. Quelques travaux restent encore à finaliser sur la sélection de variables. Concernant le cas multiclasse, nous avons proposé une stratégie mais elle ne concerne pas l'algorithme traitant le problème de manière globale [13]. Une extension de notre méthode à ce cas là doit donc être traitée. Par ailleurs, le parcours de l'espace des variables (la suppression séquentielle) est loin d'être optimal et une étude concernant ce point doit être menée. Enfin, une analyse plus approfondie de l'algorithme ainsi qu'une comparaison avec d'autres méthodes devra être également réalisée afin de mieux comprendre le comportement de l'algorithme. Quant au problème de détection de piétons, l'application de la sélection de variables a montré que les caractéristiques les plus pertinentes pour la discrimination semblent être les contours du piétons. Cela suggère donc de développer une stratégie tenant compte de ces contours. Bien que, dans cet article, l'application ciblée soit un problème de détection de piétons, il est important de souligner que la méthodologie décrite dans cet article reste générale et peut être appliquée à tout problème de sélection de variables. Des exemples d'applications sur des problèmes de bio-informatiques sont décrites en [15].

Références

- [1] R. Battiti. Using mutual information for selecting features in supervised neural net learning. *IEEE Trans. on Neural Networks*, 5(4):537–550, 1994.
- [2] L. Bottou, C. Cortes, J. Denker, H.s Drucker, I. Guyon, L. Jackel, Y. Le Cun, U. Müller, E. Säckinger, P. Simard, and V. Vapnik. Comparison of classifier methods: A case study in handwritten digit recognition. In *Proceedings of the 13th International Conference on Pattern Recognition*, 1994.
- [3] O. Chapelle, V. Vapnik, O. Bousquet, and S. Mukerjee. Choosing multiple parameters for svm. *Machine Learning*, 46(1-3):131–159, 2002.
- [4] C. Couvreur and Y. Bresler. On the optimality of the backward greedy algorithm for the subset selection problem. *SIAM Journal on Matrix Analysis and Applications*, 21(3):797–808, 2000.
- [5] N. Cristianini and J. Shawe-Taylor. *Introduction to Support Vector Machines*. Cambridge Univeristy Press, 2000.
- [6] J. Friedman. Another approach to polychotomous classification. Technical report, Department of Statistics, Stanford University, 1996.
- [7] I. Guyon and A. Elisseeff. An introduction to variable and feature selection. *Journal of Machine Learning Research*, 3:1157–1182, 2003.
- [8] I. Guyon, J. Weston, S. Barnhill, and V. Vapnik. Gene selection for cancer classification using support vector machines. *Machine Learning*, 2000.
- [9] C.-W. Hsu and C.-J. Lin. A comparison of methods for multi-class support vector machines. *IEEE Transactions on Neural Networks*, 13:415–425, 2002.
- [10] R. Kohavi and G. John. Wrappers for feature subset selection. *Artificial Intelligence (issue on Relevance)*, (97):273–324, 1997.
- [11] Y. LeCun, J.S. Denker, and S.A. Solla. *Optimal Brain Damage*, volume Neural Information Systems 2, pages 598–605. Morgan-kauffman edition, 1990.
- [12] J. Mercer. Functions of positive and negative type and their connection with the theory of integral equations. *Philos. Trans. Roy. Soc. London*, 209:415–446., 1909.
- [13] Y. Guermur and A. Elisseeff and H. Paugam-Moisy. A new multi-class SVM based on a uniform convergence result. In *International Joint Conference on Neural Networks*, volume IV, pages 183–188, 2000.
- [14] C. Papageorgiou and T. Poggio. Trainable pedestrian detection. In *Proceedings of the 1999 International Conference on Image Processing*, pages 35–39, 1999.
- [15] A. Rakotomamonjy. Variable selection using svm criteria. *Journal of Machine Learning Research*, 3:1357–1370, 2003.
- [16] B. Scholkopf and A. Smola. *Learning with Kernels*. MIT Press, 2001.
- [17] D. Skalak. prototype and feature selection by sampling and random mutation hill climbing algorithms. Proceedings of the eleventh International Conference in Machine Learning, Morgan Kaufmann, 1994.
- [18] V. Vapnik. *Statistical Learning Theory*. Wiley, 1998.
- [19] J. Weston, S. Mukherjee, O. Chapelle, M. Pontil, T. Poggio, and V. Vapnik. Feature selection for svms. In *Advances in Neural Information Processing Systems 13*, 2001.
- [20] J. Weston and C. Watkins. Multiclass support vector machines. In *Proceedings of ESANN99, Brussels*. D. Facto Press, 1999.
- [21] J. Yang and V. Honavar. Feature subset selection using a genetic algorithm. *IEEE Intelligent Systems*, 13(2):44–49, 1998.