

# Schéma d'apprentissage basé sur la sélection automatique d'exemples pertinents

G. Lebrun<sup>1</sup>

C. Charrier<sup>1</sup>

O. Lezoray<sup>1</sup>

<sup>1</sup> LUSAC EA 2607,

groupe Vision et Analyse d'Image

120 Rue de l'exode, F-50000 Saint-Lô, France

{gilles.lebrun, christophe.charrier, olivier.lezoray}@unicaen.fr

## Résumé

*La construction de processus décisionnel avec des temps de réponse courts est capital pour la réalisation d'applications ayant des contraintes de temps réel élevées. De plus, les bases d'apprentissage servant à la réalisation de ces processus contiennent généralement des exemples redondants, plus ou moins bruitées et/ou mal étiquetés. Dans cet article nous proposons un méthode de sélection des exemples les plus informatifs afin de réduire la complexité des fonctions de décision produites par les SVM. Cette méthode exploite la définition de plusieurs critères de pertinence d'exemple et la règle du Plus Proche Voisin (PPV). Les résultats expérimentaux montre que l'apprentissage des SVM avec un sous ensemble réduit d'exemples pertinents produit des fonctions de complexité réduite qui ont dans la majorité des cas des capacités en généralisation améliorées.*

## Mots clefs

Apprentissage, Classification, Quantification, SVM, PPV

## 1 Introduction

Les Séparateurs à Vaste Marge (SVM) sont des outils très performants pour la reconnaissance des formes [1]. Cependant la complexité de la fonction de décision produite par les SVM augmente linéairement avec le nombre d'exemples utilisés lors de la phase d'apprentissage [2, 3, 4, 5]. La classification de flux de données continus peut ainsi devenir rapidement problématique. Réduire la complexité des fonctions de décision produites par les SVM est devenu un enjeu essentiel [2, 6, 5, 7]. En examinant plus en détail le principe des SVM, il est possible de mettre en évidence que l'optimisation globale du problème quadratique correspondant au SVM conduit à incorporer dans la fonction de décision les exemples de la base d'apprentissage qui seront les plus difficiles à correctement classer [8, 5]. Ces exemples sont majoritairement ceux dont les données les caractérisant sont fortement bruitées et/ou mal étiquetés. Ces exemples vont croître linéairement avec la taille de la base d'apprentissage des SVM si aucune méthode n'est définie en aval pour les écarter au

préalable. Nous proposons dans cet article une méthode pour sélectionner le sous-ensemble des exemples les plus pertinents à conserver lors de la phase d'apprentissage des SVM. La méthode proposée a pour origine des travaux qui mettent en évidence que l'exploitation d'un nombre réduit d'exemples dans la règle PPV correspond à définir un principe d'inférence basé sur la minimisation du risque structural [9]. De plus la mise en œuvre de ce principe correspond à définir un schéma de compression efficace des fonctions de décision produite par les SVM. La méthode de sélection des exemples pertinents proposée permet de définir un schéma d'apprentissage en cascade en deux étapes :

1. Sélection du sous-ensemble d'exemples pertinents à partir d'un processus itératif ajoutant l'exemple le plus pertinent à chaque itération. Cet exemple est caractérisé par un critère de pertinence relative qui prend en compte l'ensemble des exemples précédemment sélectionnés, les PPV de cet exemple et la valeur d'un critère de pertinence globale pour cet exemple. Deux méthodes sont proposées pour déterminer la valeur de la pertinence globale d'un exemple à partir de la connaissance des PPV (dans la base de données initiale) de cet exemple. Le processus itératif est répété jusqu'à la satisfaction d'un critère d'arrêt qui permet d'exclure les exemples les moins significatifs et qui fixe par la même le nombre d'exemples pertinents à sélectionner.
2. Sélection des hyper-paramètres des SVM grâce à une procédure de validation croisée classique [10]. La base d'entraînement des SVM étant limitée à l'ensemble des exemples pertinents retourné à l'étape 1.

## 2 Séparateurs à Vaste Marge (SVM)

La théorie de l'apprentissage statistique de Vapnik [11] a conduit au développement d'une classe d'algorithmes désignée par l'acronyme SVM. Une des originalités de la méthode est de produire une fonction de décision binaire qui n'utilise qu'un sous-ensemble de la base d'apprentissage à partir de la minimisation d'un problème quadratique sous contraintes [11]. Les exemples appartenant à ce sous-ensemble sont nommés : vecteurs de support.

Soit  $Z^A = \{z_1, \dots, z_m\} = \{(\mathbf{x}_1, y_1), \dots, (\mathbf{x}_m, y_m)\}$  une base d'apprentissage composé de  $m$  couples (vecteur d'attributs, classe) avec  $x_i \in \mathcal{R}^n$  le vecteur d'attributs,  $y_i \in \{-1, +1\}$  la classe et  $z_i = (x_i, y_i)$  la concaténation des deux types d'information caractérisant l'exemple numéro  $i$ . Le problème quadratique à optimiser est le suivant :

$$\mathcal{W}(\alpha) = \frac{1}{2} \sum_{i,j=1}^m \alpha_i \alpha_j y_i y_j K(\mathbf{x}_i, \mathbf{x}_j) - \sum_{i=1}^m \alpha_i \quad (1)$$

sous les contraintes  $\forall i \in [1, \dots, m] : \sum_{i=1}^m y_i \alpha_i = 0, 0 \leq \alpha_i \leq C$ . La solution optimale  $\alpha^*$  définit l'ensemble  $V_S$  des vecteurs de support ( $\forall i \in [1, \dots, m] : i \in V_S \leftrightarrow \alpha_i > 0$ ) intervenant dans la fonction de décision [11] :

$$f(\mathbf{x}) = \text{sign} \left( \sum_{i \in V_S} \alpha_i^* y_i K(\mathbf{x}_i, \mathbf{x}) + b^* \right) \quad (2)$$

Le compromis entre erreur de classification et complexité du modèle dépend du choix du paramètre  $C$ . Un algorithme efficace SMO [12] a été proposé par Platt pour résoudre le problème dual. Les SVM étant des classificateurs binaires, la résolution d'un problème multiclassés est réalisée en le transformant en une combinaison de problèmes binaires [13].

### 3 Ensemble d'exemples pertinents

L'objectif est de définir une méthode qui ne conserve que les exemples pertinents présents dans une base donnée. Ces exemples pertinents sont ensuite utilisés pour réaliser la construction d'une fonction de décision efficace avec l'algorithme des SVM. La pertinence d'un ensemble d'exemples est mesurée par l'efficacité de cet ensemble à être le plus cohérent possible avec la base initiale d'exemples tout en limitant sa taille. La règle PPV est utilisée pour mesurer cette cohérence. Si l'on réalise un parallèle avec la sélection d'attributs [14, 15], la règle PPV est une approche de type *filtre* relativement aux SVM pour la sélection des exemples pertinents. Le principe d'inférence mis en œuvre correspond à la construction d'un schéma de compression efficace basé sur la minimisation du risque structurel [9, 8].

#### 3.1 Sélection des exemples pertinents

Comme la notion de pertinence individuelle d'un exemple est multiple et dépendante des autres exemples connus, plusieurs critères mesurant l'importance de la pertinence d'un exemple relativement à une base d'exemples sont utilisés [5]. Mais, avant d'aborder la définition de ces critères, commençons par donner le synopsis général de notre méthode pour la constitution de cet ensemble. Notons  $Z_m^E$  une base initiale d'exemples qui comporte  $m$  exemples et  $Z_i^P$  la base qui contient les  $i$  exemples les plus pertinents. Supposons qu'une mesure de pertinence relative  $q_r$  permet de choisir le  $i+1$ ème exemple le plus pertinent dans  $Z_m^E$  sachant que  $Z_i^P$  contient les  $i$  précédents. Le synopsis général de notre méthode est le suivant :

1.  $Z_0^P = \emptyset$  et  $i = 0$ ,
2. Sélectionner l'exemple  $j$  dont la valeur de  $q_r(z_j \in Z_m^E | Z_i^P)$  est maximale,
3. Ajouter  $z_j$  à  $Z_i^P$  pour créer  $Z_{i+1}^P$  et incrémenter  $i$ ,
4. Répéter les étapes 2 et 3 jusqu'à ce que la règle PPV avec  $Z_i^P$  soit cohérente avec  $Z_m^E$ .
5. Retourner  $Z^P = Z_i^P$

Le schéma de compression correspondant est d'autant plus performant que la taille de  $Z^P$  finale est faible. Il garantit par la même occasion une borne supérieure efficace sur l'erreur en généralisation [9]. L'efficacité de la méthode dépend de la définition de  $q_r$  et de sa relation avec  $Z_i^P$ . La base  $Z_i^P$  retournée à l'étape 5 est par la suite utilisée comme base d'apprentissage ( $Z^A$ ) pour les SVM (cf. section 2).

#### 3.2 Critère de pertinence relative

La définition de notre critère de pertinence relative  $q_r$  à partir de la règle PPV utilise deux notions qui sont :

1. Tout exemple qui est dans  $Z_m^E$  et qui est correctement classé par la règle PPV exploitant  $Z_i^P$  n'est pas pertinent, car il est redondant.
2. Il est possible de définir une mesure globale de pertinence  $q_g$  qui caractérise la pertinence d'un exemple  $z$  à partir de la connaissance de  $Z_m^E$ .

Ces deux notions autorisent la remise en cause de la pertinence globale d'un exemple s'il est redondant avec l'ensemble des exemples les plus pertinents (relativement à cette mesure globale) déjà sélectionnés. A partir de la notion n°1, nous définissons le sous-ensemble d'exemples pertinents candidats  $Z_i^c$  (le  $c$  en exposant étant une contraction de candidat) comme :

$$Z_i^c = \{z | z = (\mathbf{x}, y) \in Z_m^E \wedge z \notin Z_i^P \wedge \mathbf{V}(\mathbf{x}, Z_i^P) \neq y\} \quad (3)$$

avec  $\mathbf{V}(\mathbf{x}, Z_i^P)$  donnant la classe du plus proche voisin de  $\mathbf{x}$  à partir de  $Z_i^P$ . La formulation de pertinence relative pour un exemple  $z$  donné à partir des deux notions précédentes est la suivante :

$$q_r(z \in Z_m^E | Z_i^P) = \begin{cases} q_g(z, Z_m^E) & \text{si } z \in Z_i^c \\ -\infty & \text{sinon} \end{cases} \quad (4)$$

#### 3.3 Critères de pertinence globale

Parmi les règles PPV utilisant un ensemble  $Z^P$  donné et qui sont cohérentes avec  $Z^E$ , celles exploitant un ensemble  $Z^P$  de taille minimale sont celles qui ont un risque structurel le plus faible [9]. De plus le fait d'avoir un nombre réduit d'exemples garantit une compression efficace de la fonction de décision produite par un SVM. Il est donc important d'ajouter à chaque étape de notre méthode l'exemple  $z$  qui permet de classer correctement un maximum d'autres exemples de la base. Nous proposons deux critères pour réaliser cette estimation de pertinence globale qui sont basés sur les résultats de différents travaux [16, 9, 17, 18].

**Critère de pertinence globale n°1.** Ce critère a pour inspiration les travaux de Shin et al [18] qui utilisent la règle  $k$ -PPV pour définir une mesure de probabilité locale de chaque classe. Cette mesure correspond à déterminer pour un exemple donné ses  $k$  plus proches voisins, puis à calculer la probabilité de chaque classe à partir de celles de ces  $k$  voisins. Cette mesure caractérise la confiance dans la classe de l'exemple considéré. En effet, si la probabilité estimée est élevée pour la classe correspondant à celle de l'exemple et que la valeur de  $k$  est grande, alors la probabilité que l'exemple ne soit pas de cette classe est faible. Pour obtenir la mesure de pertinence globale  $q_{g1}$ , nous considérons qu'un exemple  $z$  est d'autant plus pertinent que le nombre  $k$  est élevé à condition que tous les  $k$  voisins les plus proches de  $z$  soient de la même classe que l'exemple considéré. La valeur maximale de  $k$  caractérise la confiance dans la classe de l'exemple et ainsi sa pertinence à être utilisé dans la règle PPV. Elle caractérise également le fait que l'exemple est dans une zone à forte densité d'exemples de la même classe.

Soit  $n\text{-PPV}_{mc}(k, z, Z)$  la fonction qui retourne le nombre de PPV de la même classe que  $z$  parmi les  $k$  PPV de  $z$  dans  $Z$  :

$$q_{g1}(z, Z) = \max_k (k \cdot I(n\text{-PPV}_{mc}(k, z, Z) = k)) \quad (5)$$

avec  $I(\text{vrai}) = 1$  et  $I(\text{faux}) = 0$ .

**Critère pertinence globale n°2.** La définition de ce critère a pour inspiration une méthode de sélection d'attributs [17] basé sur l'exploitation d'une marge d'hypothèse  $\gamma_h(z)$ . Elle mesure l'efficacité de la règle PPV pour un exemple  $z$  donné à partir de la base d'exemples et a été définie [16] comme étant égale à

$$\gamma_h(z) = \gamma_h(\mathbf{x}, y) = \frac{1}{2} (\|\mathbf{x}_i - \mathbf{x}\| - \|\mathbf{x}_j - \mathbf{x}\|) \quad (6)$$

avec  $\mathbf{x}_i$  et  $\mathbf{x}_j$  les vecteurs caractérisant respectivement l'exemple  $z_i$  le plus proche  $z$  avec une classe différente de celle de  $z$  et l'exemple  $z_j$  le plus proche  $z$  avec la même classe que  $z$ . Dans cette méthode de sélection d'attributs, l'idée est qu'un attribut est d'autant plus important que sont exploitation à partir d'une règle PPV est efficace. Pour noter cette importance, un ensemble d'exemples ( $z_k$ ) sont choisis aléatoirement et les valeurs correspondantes des  $\gamma_h(z_k^l)$  sont déterminées pour chacun des attributs ( $z_k^l$  désigne l'attribut  $l$  de l'exemple  $k$ ). L'importance de l'attribut  $l$  est ensuite calculée à partir des valeurs  $\gamma_h(z_k^l)$  correspondantes.

Nous reprenons ce principe sans définir une procédure stochastique et le transposons à la notion d'exemple pertinent. L'idée directrice pour la définition de ce nouveau critère est que les exemples qui permettent de réaliser une classification correcte à partir de la règle PPV ont des valeurs de marge d'hypothèse importantes et ceux conduisant à produire des erreurs ont des valeurs de marge faibles ou négatives. Nous définissons deux propriétés qui sont utiles

à la sélection des exemples  $z'$  ayant comme PPV l'exemple  $z$  suivant la classe de  $z$  et  $z'$  :

1. Le voisin (V) le plus proche et de la même classe (m) d'un exemple  $z = (\mathbf{x}, y)$  donné :

$$V_m(\mathbf{x}, y) = \arg \min_{\substack{(\mathbf{x}', y') \in Z \\ y' = y \wedge \mathbf{x}' \neq \mathbf{x}}} (d(\mathbf{x}, \mathbf{x}')) \quad (7)$$

2. Le voisin le plus proche et d'une classe différente (d) d'un exemple  $z = (\mathbf{x}, y)$  donné :

$$V_d(\mathbf{x}, y) = \arg \min_{\substack{(\mathbf{x}', y') \in Z \\ y' \neq y}} (d(\mathbf{x}, \mathbf{x}')) \quad (8)$$

A partir de ces deux propriétés, nous définissons l'ensemble des marges d'hypothèse  $\gamma_h(z')$  relatives à un exemple  $z$  :

$$\Gamma_h(z) = \{\gamma_h(z') : V_m(z') = z \vee V_d(z') = z\} \quad (9)$$

Un exemple  $z$  de la base  $Z^E$  est d'autant plus pertinent que les marges d'hypothèse dans  $\Gamma_h(z)$  sont élevées. Plus les valeurs dans  $\Gamma_h(z)$  seront élevées et positives, et plus cet exemple sera un bon candidat pour son exploitation par la règle PPV. Nous avons choisi la valeur moyenne des marges de  $\Gamma_h(z)$  comme mesure de la pertinence d'un exemple :

$$q_{g2}(z, Z) = \overline{\Gamma_h(z)} \quad (10)$$

avec  $\overline{\Gamma_h(z)}$  désignant la valeur moyenne de  $\Gamma_h(z)$  (par convention la moyenne d'un ensemble vide est égale 0).

### 3.4 Critère d'arrêt prématuré

Dans le synopsis général (c.f. section 3.1) le critère d'arrêt de l'étape 4 impose la cohérence avec  $Z^E$ , imposant par la même que la règle PPV avec l'ensemble  $Z^P$  sélectionné ne fasse aucune erreur de prédiction dans  $Z^E$ . Cependant, lorsque les exemples sont fortement bruités ou que des erreurs de labélisation se sont produites lors de la constitution de  $Z^E$ , il n'est plus souhaitable d'atteindre cette cohérence. En effet, l'ajout d'exemples non pertinents (*i.e.* de faibles pertinences suivant les critères  $q_{g1}$  ou  $q_{g2}$ ) peut-être contre productif. L'idée est alors d'exploiter ces critères de pertinences pour produire un arrêt prématuré (*i.e.* avant cohérence) afin de produire une base d'apprentissage  $Z^A$  de taille plus réduite tout en garantissant un pouvoir de généralisation élève à partir d'elle.

**Critère d'arrêt prématuré n°1.** Lorsque le critère  $q_{g1}$  est utilisé dans la définition de  $q_r$ , les exemples sont considérés comme non-pertinents lorsque le nombre de voisins de la même classe devient trop faible. La règle d'arrêt prématuré choisie est :

$$(\forall z \in Z_i^c : q_{g1}(z, Z^E) < 2) \rightarrow \text{Arrêt} \quad (11)$$

**Critère d'arrêt prématuré n°2.** Lorsque le critère  $q_{g2}$  est utilisé dans la définition de  $q_r$ , les exemples sont considérés comme non-pertinents lorsque la moyenne des marges d'hypothèse  $\gamma_h$  n'est plus positive. La règle d'arrêt prématuré choisie est :

$$(\forall z \in Z_i^c : q_{g2}(z, Z^E) \leq 0) \rightarrow \text{Arrêt} \quad (12)$$

## 4 Schéma d'apprentissage

Nous avons défini un schéma d'apprentissage en cascade qui a pour but de réaliser la sélection d'un modèle performant pour la production de fonctions de décision de complexité réduite avec des SVM. La figure 1 illustre les deux étapes principales de la sélection de modèle en cascade relative à ce schéma de simplification. La première étape de ce schéma réalise la sélection d'un sous-ensemble d'exemples pertinents comme décrit en section 3 (avec exploitation d'un critère d'arrêt prématuré). La seconde étape réalise la recherche des hyper-paramètres optimaux lorsque les SVM sont entraînés à partir de l'ensemble d'exemples pertinents sélectionnés par la première étape.

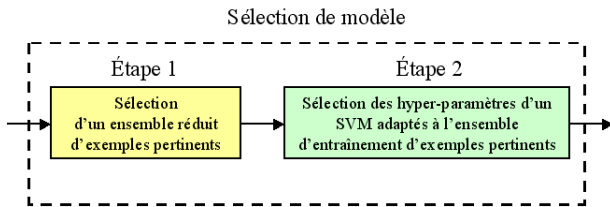


Figure 1 – Synopsis des deux étapes de la sélection de modèle en cascade.

L'exploitation des deux critères de pertinence  $q_{g1}$  et  $q_{g2}$  n'a pas permis de mettre en évidence la supériorité d'un critère par rapport à l'autre. En effet, lorsque l'on compare les taux d'erreur des deux fonctions de décision produites à partir de ces critères suivant la base d'exemples utilisée l'un est plus performant que l'autre ou réciproquement [5]. Le schéma d'apprentissage produit donc deux bases d'apprentissage  $Z_1^A$  et  $Z_2^A$  à l'étape 1. Pour l'étape 2, un noyau gaussien  $K(\mathbf{x}_i, \mathbf{x}_j) = \exp\left(-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|^2}{2\sigma^2}\right)$  est utilisé pour les SVM. Pour chaque base d'apprentissage simplifiée  $Z^A$  produite, une recherche des hyper-paramètres optimaux  $(C, \sigma)$  pour un SVM (ou ensemble de SVM si le problème est multi-classe) est réalisée à partir d'une procédure de validation croisée en 10 parties. La base  $Z^A$  qui permet d'obtenir l'erreur de validation croisée la plus faible est sélectionnée, ainsi que les hyper-paramètres correspondants. Lorsque l'écart est faible entre les estimations de l'erreur de validation croisée (inférieur à 1% en écart relatif) avec les deux critères  $c_{g1}$  et  $c_{g2}$ , c'est la base simplifiée de taille la plus faible qui est sélectionnée. Ceci afin de favoriser la réduction de la complexité de la fonction de décision produite. La base simplifiée  $Z^A$  et les hyper-paramètres  $(C, \sigma)$  sélectionnés sont utilisés pour

produire, à partir de l'algorithme des SVM, la fonction de décision simplifiée. Lorsque la base d'apprentissage comporte plus de deux classes, c'est un schéma de combinaison *un contre un* [13] qui est utilisé pour produire la fonction de décision simplifiée et chaque SVM binaire intervenant dans ce schéma de combinaison utilise les mêmes hyper-paramètres  $(C, \sigma)$ .

## 5 Expérimentations

Nous avons appliqué notre schéma de simplification en cascade à plusieurs bases de données. La base QIjpeg2M correspond à une étude de qualité visuelle d'images couleur pour différents niveaux de compression avec JPEG2000 [19]. Toutes les autres bases proviennent de l'*UCI repository* [20] et sont régulièrement utilisés dans des *benchmarks* d'apprentissage. Pour mettre en évidence l'efficacité de notre méthode de sélection d'exemples pertinents, nous avons comparé un apprentissage classique (*i.e.* la sélection dans le schéma 1 retourne tous les exemples comme pertinents) à notre schéma d'apprentissage. Le tableau 1 illustre les résultats obtenus avec ces deux procédures d'apprentissage. Les noms des bases ayant un suffixe '-2c' correspondent à une transformation en problèmes binaires des bases initiales (sans le suffixe) comportant plus de deux classes [5]. La raison principale de ce type de transformation est que les SVM sont des classificateurs binaires. L'influence de la sélection d'un schéma de combinaison par rapport à un autre n'est pas une opération neutre [13] et la mesure de l'effet plus ou moins positif de techniques de simplification pour les SVM sur des problèmes multi-classes est donc dépendante du schéma choisi. Pour des problèmes binaires seul l'effet de la simplification est pris en compte (*ex.* Glass et Glass-2c).

Les résultats dans tableau 1 montrent que dans la majorité des cas l'apprentissage des SVM avec un sous ensemble d'exemples pertinents permet d'obtenir une fonction de décision qui a des meilleures capacités en généralisation. Pour certaines bases l'amélioration peut-être très significative (*ex.* Qijpeg2M et Bupa), illustrant par la même l'importance de supprimer les exemples non-significatifs grâce à l'exploitation d'un critère d'arrêt prématuré. La réduction du nombre d'exemples utilisés par les fonctions de décision produites est importante (en particulier pour les bases Qijpeg2M, Iris, Glass, Segment et Bupa). Remarquons que le fait d'avoir un maximum d'exemples en apprentissage avec les SVM pour réduire l'écart entre l'erreur empirique et l'erreur réelle doit être contrebalancé par la pertinence des exemples exploités lors de l'apprentissage. La colonne  $q_g$  dans le tableau 1 illustre également qu'il n'est pas possible de choisir un des deux critères  $q_{g1}$  ou  $q_{g2}$  comme étant supérieur à l'autre.

## 6 Conclusion

Dans cet article nous avons proposé une méthode permettant de sélectionner un sous-ensemble d'exemples pertinents à partir de la définition de critères de pertinence

Base	classique		schéma en cascade		
	$e_V$	VS %	$q_g$	$e_V$	VS %
Qljpeg2M	22%	60%	1	<b>17,1%</b>	<b>8,6%</b>
Qljpeg2M-2c	0,9%	13%	1	<b>0%</b>	<b>1,7%</b>
Iris	0%	22%	1	0%	<b>4,2%</b>
Iris-2c	0%	3,3%	1, 2	0,0%	<b>1,7%</b>
Wine	0%	28%	1	0%	<b>8,3%</b>
Wine-2c	0%	19%	2	0%	<b>7,6%</b>
Glass	21%	72%	1	23,1%	<b>8,6%</b>
Glass-2c	13%	40%	1	<b>10,3%</b>	<b>5,7%</b>
Segment	2,9%	58%	1	<b>2,8%</b>	<b>7,4%</b>
Segment-2c	8,6%	24%	2	<b>5,7%</b>	<b>19,4%</b>
Bupa	31%	62%	2	<b>23,9%</b>	<b>24,5%</b>
Ionosphere	4,3%	33%	2	7,25%	<b>17,1%</b>
Breast-cancer	2,2%	11%	1	<b>1,48%</b>	<b>2,0%</b>
OptDigits-2c	0%	6%	1, 2	0%	<b>1,8%</b>
Pima	23%	54%	2	<b>21,7%</b>	<b>3,1%</b>
Vehicle	17%	59%	2	<b>16,7%</b>	<b>20,6%</b>
Vehicle-2c	1,8%	20%	2	<b>0,6%</b>	<b>8,8%</b>
Letter-2c	0,97%	4,3%	2	<b>0,32%</b>	<b>2,6%</b>
Pendigits-2c	1,2%	1,5%	2	<b>0,27%</b>	<b>0,83%</b>

Tableau 1 – Comparaison entre un apprentissage classique avec des SVM et notre schéma en cascade. Les colonnes  $e_V$  donnent le taux d'erreur mesuré à partir d'une même base de validation. Les colonnes VS% donnent le pourcentage d'exemples utilisés comme vecteurs de support dans la fonction de décision produite par rapport à ceux dans la base initiale  $Z^E$ . La colonne  $q_g$  indique le numéro du critère de pertinence globale utilisé pour constituer la base d'apprentissage des SVM (les deux numéros sont indiqués si elles sont identiques).

basés sur l'exploitation de la règle PPV. Les résultats expérimentaux mettent en évidence que cette sélection permet de réduire la complexité des processus décisionnels produit par les SVM tout en garantissant de bonnes performances en généralisation. Motivé par les résultats obtenus avec les bases '2c' (cf. tableau 1), nos travaux futurs porteront sur l'analyse comparative de l'effet d'une sélection d'un sous-ensemble d'exemples pertinents pour chacun des sous-problèmes binaires à celui d'une sélection globale au problème multi-classes dans le cadre de l'exploitation de différents schémas de combinaison binaire couramment utilisés avec les SVM.

## Références

- [1] S. Abe. *Support Vector Machines for Pattern Classification*. Springer, 2005.
- [2] G. Lebrun, C. Charrier, et O. Lezoray. Réduction du temps d'apprentissage des svm par quantification vectorielle. Dans *CORESA*, pages 223–226, 2004.
- [3] G. Lebrun, C. Charrier, O. Lezoray, C. Meurie, et H. Cardot. Classification rapide de pixels par svm dans un espace couleur hybride. Dans *CORESA*, pages 231–236, 2005.
- [4] I. Steinwart. Sparseness of support vector machines - some asymptotically sharp bounds. Dans *NIPS*, pages 169–184, 2004.
- [5] G. Lebrun. *Sélection de modèles pour la classification supervisée avec des SVM (Séparateurs à Vaste Marge). Application en traitement et analyse d'images*. Thèse de doctorat, Université de Caen, Novembre 2006.
- [6] I. W. Tsang, J. T. Kwok, et P.-M. Cheung. Core vector machines : Fast SVM training on very large data sets. *JMLR*, 6 :363–392, 2005.
- [7] S. S. Keerthi, O. Chapelle, et D. Decoste. Building support vector machines with reduced classifier complexity. *Journal of Machine Learning Research*, 7 :1493–1515, 2006.
- [8] B. Karacali, R. Ramanath, et W. E. Snyder. A comparative analysis of structural risk minimization by support vector machines and nearest neighbor rule. *Pattern Recognition Letters*, 25(1) :63–71, 2004.
- [9] B. Karacali et H. Krim. Fast minimization of structural risk by nearest neighbor rule. *IEEE Transactions on Neural Networks*, 14(1) :127–137, 2003.
- [10] G. Lebrun, O. Lezoray, C. Charrier, et H. Cardot. An ea multi-model selection for svm multiclass schemes. Dans *IWANN*, 2007, A paraître.
- [11] V. N. Vapnik. *Statistical Learning Theory*. New York, wiley édition, 1998.
- [12] J. Platt. Fast training of support vector machines using sequential minimal optimization, advances in kernel methods-support vector learning. Dans *MIT Press*, pages 185–208, 1999.
- [13] C.-W. Hsu et C.-J. Lin. A comparison of methods for multiclass support vector machines. Dans *IEEE Transactions in Neural Networks*, volume 13, pages 415–425, 2002.
- [14] M. Kudo et J. Sklansky. Comparison of algorithms that select features for pattern classifiers. Dans *Pattern Recognition*, volume 33, pages 25–41, 2000.
- [15] I. Tsamardinos et C. Aliferis. Towards principled feature selection : Relevancy, 2003.
- [16] Koby Crammer, Ran Gilad-Bachrach, Amir Navot, et Naftali Tishby. Margin analysis of the lqv algorithm. Dans *NIPS*, pages 462–469, 2002.
- [17] Ran Gilad-Bachrach, Amir Navot, et Naftali Tishby. Margin based feature selection - theory and algorithms. Dans *ICML*, 2004.
- [18] Hyunjung Shin et Sungzoon Cho. Pattern selection for support vector classifiers. Dans *IDEAL*, pages 469–474, 2002.
- [19] Laboratory for Image & Video Engineering, University of Texas (Austin). LIVE Image Quality Assessment Database. <http://live.ece.utexas.edu/research/Quality>, 2002.
- [20] C. Blake et C. Merz. Uci repository of machine learning databases. advances in kernel methods, support vector learning., 1998.