Décisions en environnement non stationnaire par méthodes d'ensembles via One-class SVM - application à la segmentation d'images texturées

Pierre Beauseroy, André Smolarz, Xiyan He

▶ **To cite this version:**

Pierre Beauseroy, André Smolarz, Xiyan He. Décisions en environnement non stationnaire par méthodes d'ensembles via One-class SVM - application à la segmentation d'images texturées. 42èmes Journées de Statistique, 2010, Marseille, France. inria-00494710

HAL Id: inria-00494710
https://inria.hal.science/inria-00494710

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.

Décisions en environnement non stationnaire par méthodes d'ensembles via One-class SVM application à la segmentation d'images texturées

Pierre Beauseroy , Andre Smolarz & Xiyan He

Institut Charles Delaunay (FRE 2848) – Université de Technologie de Troyes, 12, Rue Marie Curie, BP 2060, 10010 Troyes Cedex
{pierre.beauseroy , andre.smolarz , xiyan.he}@utt.fr

Résumé : *Nous proposons une méthode de décision basée sur un ensemble d'espaces de représentation, dans le but d'optimiser ou de préserver les performances d'un système décisionnel en présence de bruit, de perte d'information ou de non-stationnarité. Sa spécificité réside dans le fait que seules les règles définies dans des espaces non perturbés par la non stationnarité sont prises en compte pour la décision finale. Nous abordons le problème au moyen d'un ensemble de One-class SVM's. Nous présentons ici quelques résultats obtenus en segmentation d'images texturées qui constitue une application tout à fait appropriée pour illustrer le problème et les performances de la méthode.*

Abstract : *We propose a decision method based on a set of representation spaces. To tackle the problem of optimizing or preserving the performances of a decision-making system in the case of perturbations due to noise, loss of information or non-stationnarity. Its specificity lies in the fact that only the rules defined in non corrupted spaces are taken into account for the final decision. We adress the problem using One-class SVM's method. Results obtained in textured image segmentation are presented. They illustrate the problem and the method performances.*

Mots clés : décision, non-stationnarité, apprentissage, méthodes d'ensembles, One-class SVM, segmentation d'images texturées.

1 Introduction

Dans le cadre de la surveillance de systèmes complexes, la mise en œuvre de méthodes de décision par apprentissage nécessite de disposer d'une base d'observations étiquetées. Généralement chaque observation est constituée d'un ensemble d'attributs prédéfinis qui sont aptes à caractériser les états du système. Chaque observation peut être considérée comme la réalisation d'une variable aléatoire qui obéit à une loi de probabilité modélisant l'état du système. L'ensemble de données est utilisé selon diverses modalités pour déterminer une règle de décision et pour estimer les performances de la règle. Cette dernière est ensuite utilisée pour attribuer un label à de nouvelles observations issues du même système ou des mêmes lois de probabilité. Les performances de la règle ne dépendent que de la qualité de l'apprentissage.

Un problème particulier se pose si les lois d'une partie des mesures conditionnellement aux mesures restantes changent. Ces évolutions peuvent venir de dysfonctionnements ou de pannes de certains capteurs, de bruits instationnaires ou du fait que tous les capteurs ne perçoivent pas le même état du système. Dans de telles situations, la règle apprise n'est plus adaptée aux observations à classer et les performances se dégradent très sensiblement.

Il est possible de faire un test préalable pour savoir si l'observation peut venir de la distribution apprise ou pas et de ne classer que les observations conformes à l'apprentissage. Cette solution rend le système de surveillance inopérant en cas de perturbations. Dans ce type de situation, pour disposer d'un système de décision qui reste opérationnel et qui limite la dégradation des performances nous proposons de prendre la décision en fusionnant les décisions individuelles d'un ensemble de règles dont chacune n'exploite qu'une fraction des mesures provenant du système à surveiller. Ceci correspond à un cas particulier des méthodes d'ensembles travaillant sur un ensemble de sous-espaces de représentation (Breiman, 2001). Dans ce cas, seules sont affectées les performances des règles qui exploitent des mesures perturbées. Nous proposons donc de concevoir un système qui écarte, si possible, les règles affectées et qui conserve les autres, de sorte que seules participent à la décision finale, les règles qui s'appuient sur une perception conforme à leur apprentissage. Ecarter les règles dont les mesures sont perturbées est un problème qui s'apparente à la détection de nouveauté ou d'anomalie. Il peut être abordé à l'aide de méthodes de classification par apprentissage qui consistent à construire un modèle de représentation des données *normales* et à l'utiliser pour détecter les données *non conformes*. Dans la littérature, ce type de problème est parfois traité comme un problème de classification à deux classes (Kassab and Alexandre, 2009) :

- Une classe cible ou classe *positive* dont les données sont disponibles.
- Une autre classe, dont les données sont difficiles ou impossibles à obtenir. Cette classe correspond à la classe d'anomalie ou classe *négative*. De nombreuses méthodes ont été proposées dans ce cadre (Li and Liu, 2003; Yu et al., 2003).

La classification à une classe est une autre approche de ce type de problèmes. Dans ce cas, l'apprentissage ne tient compte que des exemples positifs. L'idée principale étant d'apprendre la classe *normale* en ajustant le modèle aux données *positives*. Les nouvelles observations sont classées en comparant leur score, donné par le modèle, à un seuil de décision. Il existe deux catégories de méthodes de ce type (Camci and Chinnam, 2008) : les méthodes statistiques (estimation de la densité de probabilité, Parzen, k-ppv, modèle de mélange de Gaussiennes...) et les méthodes reposant sur l'optimisation d'une fonction de décision au sein d'une classe de fonctions comme par exemple les réseaux de neurones ou des approches basées sur les SVM. La méthode de *One-class SVM*, proposée par B. Schölkopf est l'une d'entre elles (Schölkopf et al., 2001; Hayton et al., 2000; Manevitz and Yousef, 2001). L'approche que nous avons adoptée est inspirée de la méthode *One-class SVM*. Dans le système décisionnel que nous proposons, chaque classifieur à une classe joue deux rôles à la fois : la validation de la règle et la décision en faveur de la classe modélisée.

Dans un premier temps, nous présentons la méthode de *One-class SVM*. Nous présentons ensuite notre approche, baptisée SEROSVM (Sélection d'Espaces de Représentation via One-class SVM), de sélection optimale d'espaces de représentation par agrégation d'un ensemble de classifieurs basés sur les *One-class SVM*. Enfin, nous présentons quelques résultats issus des études expérimentales que nous avons menées en segmentation d'images texturées.

2 One-class SVM

(Schölkopf et al., 2001) formule ainsi le problème de la classification à une classe :
En considérant un ensemble de données obéissant à une loi de probabilité P, la question

est d'estimer un sous-ensemble S de l'espace de représentation des données tel que la probabilité qu'une réalisation tirée de la loi P appartienne à $\bar{S}$ (complément de S) soit bornée par un seuil spécifié a priori.

L'objectif de la méthode est d'estimer une fonction f positive sur S et négative sur $\bar{S}$. Soit $\mathcal{X}^m \subseteq \mathbb{R}^m$ un espace de représentation de dimension m et $\mathcal{A}_n = \{\boldsymbol{x}_i \in \mathcal{X}^m \mid i = 1, \ldots, n\}$ un ensemble d'apprentissage de taille n. Soit ϕ une transformation non-linéaire dont le produit scalaire de son image puisse être calculé par un noyau simple $K(\boldsymbol{x}_i, \boldsymbol{x}_j) = \langle \phi(\boldsymbol{x}_i), \phi(\boldsymbol{x}_j) \rangle$, comme par exemple un noyau gaussien $K_\sigma(\boldsymbol{x}_i, \boldsymbol{x}_j) = e^{-\frac{\|\boldsymbol{x}_i - \boldsymbol{x}_j\|^2}{2\sigma^2}}$.

Il s'agit alors de projeter les données dans un espace induit par le noyau utilisé puis d'en séparer une proportion $1-\nu$ de l'origine par l'hyperplan le plus éloigné de l'origine – on parle de l'hyperplan de marge maximale. Dans l'espace initial, l'hyperplan séparateur correspond à l'enveloppe de la région de petit volume qui contient une proportion (fonction de ν) des données d'apprentissage. Pour déterminer cet hyperplan, il faut trouver $\mathbf{w}$ le vecteur normal et un seuil ρ, solutions du problème suivant :

$$\begin{cases} \min_{\mathbf{w}, \boldsymbol{\xi}, \rho} & \frac{1}{2}\|\mathbf{w}\|^2 + \frac{1}{\nu n}\sum_{i=1}^{n} \xi_i - \rho \\ \text{sous les contraintes} & \langle \mathbf{w}, \phi(\boldsymbol{x}_i) \rangle \geq \rho - \xi_i, \quad \xi_i \geq 0 \end{cases} \tag{1}$$

Des variables ressorts ξ_i, associées à chaque observation, sont introduites pour pouvoir relâcher la contrainte pour certains exemples de l'ensemble d'apprentissage. L'algorithme d'optimisation consiste à trouver un compromis entre la maximisation de la marge et la minimisation du ressort moyen. Les $\boldsymbol{x}_i \in \mathcal{A}_n$ qui se situent du mauvais côté de l'hyperplan séparateur sont classées comme des anomalies et la distance entre une anomalie et l'hyperplan vaut $\xi_i / \|\mathbf{w}\|$ avec $\xi_i > 0$. La distance entre la marge et l'origine est égale à $\rho / \|\mathbf{w}\|$ et le paramètre $\nu \in [0, 1]$ est une borne supérieure du taux d'anomalies, mais aussi une borne inférieure du taux de vecteurs supports.

Le problème dual s'écrit : $\begin{cases} \min_{\boldsymbol{\alpha}} & \frac{1}{2}\sum_{i,j=1}^{n} \alpha_i \alpha_j K(\boldsymbol{x}_i, \boldsymbol{x}_j) \\ \text{sous les contraintes} & 0 \leq \alpha_i \leq \frac{1}{\nu n}, \quad \sum_{i=1}^{n} \alpha_i = 1 \end{cases}$ (2)

et la fonction de décision est donnée par : $f(\boldsymbol{x}) = sgn\left(\sum_{i=1}^{n} \alpha_i K(\boldsymbol{x}_i, \boldsymbol{x}) - \rho\right)$ (3)

En résumé, on peut donc distinguer trois types d'observations d'apprentissage :
- les anomalies pour lesquelles on a : $f(\boldsymbol{x}_i) < 0 \quad \alpha_i = 1/\nu n \quad \xi_i > 0$
- Les observations normales pour lesquelles on a : $f(\boldsymbol{x}_i) > 0 \quad \alpha_i = 0 \quad \xi_i = 0$
- Les vecteurs de support pour lesquels on a : $f(\boldsymbol{x}_i) = 0 \quad 0 < \alpha_i < \frac{1}{\nu n}$

3 Approche proposée - SEROSVM

Soit donc $\boldsymbol{X} \in \mathcal{X}^m$ le vecteur aléatoire d'attributs. Notre méthode repose sur le fait qu'en situation d'environnement *normal* pour le système sous surveillance, l'ensemble des m attributs (variables aléatoires) obéissent conjointement au modèle décrivant la classe *normale*. On dira dans ce cas que le modèle ou que l'espace $\mathcal{X}^m$ est *homogène*. Dans le cas contraire (environnement perturbé), différentes situations peuvent se présenter pour lesquelles, seule une partie des composantes de $\boldsymbol{X}$ seront aptes à décrire la classe *normale*. Le ou les sous-espaces définis sur cette partie des attributs constitueront alors des sous-espaces *homogènes*. Plus précisément, les classifieurs étant construits dans des sous-espaces

de représentation extraits de l'espace initial, nous faisons l'hypothèse que ceux qui sont construits sur des sous-espaces *homogènes* doivent être plus pertinents et plus fiables pour classer des observations provenant d'un environnement perturbé.

L'approche SEROSVM que nous proposons consiste donc à générer dans un premier temps un ensemble $\left\{ \mathcal{X}_\ell^d \subset \mathcal{X}^m \mid \ell = 1, \ldots, L \right\}$ de L sous-espaces de représentation en effectuant L tirages aléatoires de d attributs parmi m (on notera $\boldsymbol{X}^{(\ell)}$, le vecteur aléatoire constitué des d composantes de $\boldsymbol{X}$ décrivant $\mathcal{X}_\ell^d$). On constitue ensuite pour chacun 2 bases d'apprentissage $\mathcal{A}_n^{\ell,1}$ et $\mathcal{A}_n^{\ell,2}$ (associées à chacune des 2 classes) au moyen desquelles on apprend 2 règles de décisions de one-class SVM : f_ℓ^1 et f_ℓ^2. On obtient alors un ensemble de classifieurs h_ℓ définis ainsi :

$$h_\ell(\boldsymbol{x}^{(\ell)}) = \mathbb{1}_{f_\ell^1(\boldsymbol{x}^{(\ell)})>0} - \mathbb{1}_{f_\ell^2(\boldsymbol{x}^{(\ell)})>0} \quad \text{avec} \quad \mathbb{1}_{f_\ell^k(\boldsymbol{x}^{(\ell)})>0} = \begin{cases} 1 & \text{si} \quad f_\ell^k(\boldsymbol{x}^{(\ell)}) > 0 \\ 0 & \text{sinon} \end{cases} \tag{4}$$

La définition (4) entraîne, selon la valeur de h_ℓ, 4 cas de décisions (Cf. tableau 1). La décision finale repose sur un vote majoritaire associé à un seuil θ estimé sur la base du taux d'erreur minimal obtenu sur un ensemble test (formule 5).

	$f_\ell^1(\boldsymbol{x}^{(\ell)}) > 0$	$f_\ell^1(\boldsymbol{x}^{(\ell)}) \leq 0$
$f_\ell^2(\boldsymbol{x}^{(\ell)}) > 0$	$h_\ell(\boldsymbol{x}^{(\ell)}) = 0$ décision ambiguë	$h_\ell(\boldsymbol{x}^{(\ell)}) = -1$ décision classe 2
$f_\ell^2(\boldsymbol{x}^{(\ell)}) \leq 0$	$h_\ell(\boldsymbol{x}^{(\ell)}) = 1$ décision classe 1	$h_\ell(\boldsymbol{x}^{(\ell)}) = 0$ $\mathcal{X}_\ell^d$ est hétérogène

TAB. 1 – Différents types de décisions selon la valeur de h_ℓ

$$y_{final} = sgn\left(\sum_\ell h_\ell(\boldsymbol{x}^{(\ell)}) + \theta \right) \tag{5}$$

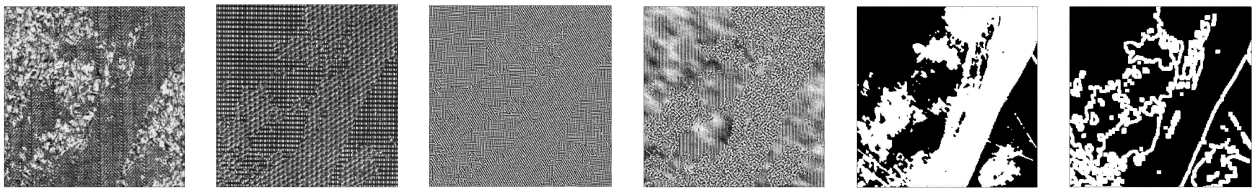
4 Expérimentations

La segmentation d'images texturées constitue un support adéquat pour illustrer les diverses situations *d'homogénéité* et *d'hétérogénéité*. Nous présentons les résultats de la méthode sur plusieurs images texturées de taille 256×256 pixels comprenant des textures de (Brodatz, 1966) ou des textures markoviennes (Smolarz, 1997) (Cf. figure 1). À chaque pixel est attachée une variable aléatoire représentant son niveau de gris. L'espace de représentation initial $\mathcal{X}^m$ est décrit par les m variables attachées aux pixels de chaque voisinage carré de $m = p \times p$ pixels de l'image. Le pixel occupant la position centrale de chaque voisinage est le pixel à classer.

Les résultats présentés ici ont été obtenus dans les conditions expérimentales suivantes :
- dimension de l'espace de représentation initial : $m = p \times p = 5 \times 5 = 25$
- dimension de chaque sous-espace de représentation : $d = 5$
- nombre total de sous-espaces de représentation extraits aléatoirement : $L = 100$
- le noyau des *One-class SVM* est un noyau gaussien $K_\sigma(\boldsymbol{x}_i, \boldsymbol{x}_j)$

Les deux paramètres ν et σ ont été fixés, pour chaque image, après une recherche de la combinaison offrant la meilleure performance sur un ensemble de test. Leurs valeurs sont précisées pour chaque image à traiter (Cf. figure 1). La figure 2 présente les résultats de

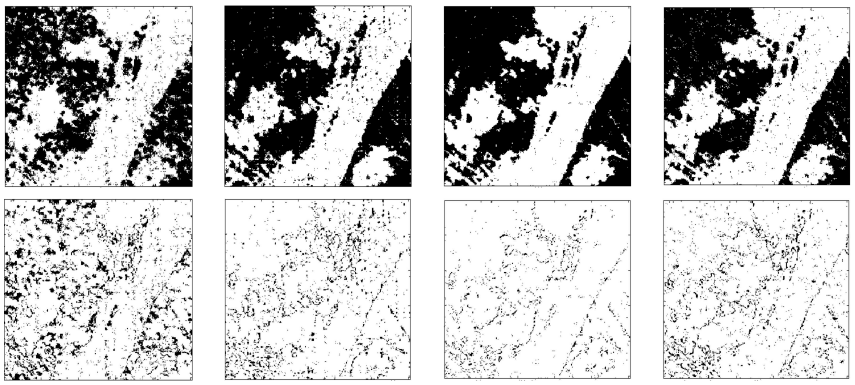
la segmentation ainsi que la distribution spatiale des erreurs. Malgré la complexité des frontières, les segmentations obtenues pour chaque image de test montrent, en comparant avec la segmentation idéale, de bonnes performances de la méthode proposée. Des comparaisons que nous avons menées avec des méthodes concurrentes (He et al., 2009) confirment ces résultats. On observe également que la plupart des erreurs de classification se trouvent dans les zones frontières qui représentent les données pour lesquelles l'espace de représentation initial est fortement hétérogène.



(0.1, 65) (0.15, 40) (0.1, 75) (0.1, 100)

a b c d e f

FIG. 1 – Images tests (a - d), carte des régions (e) et carte des frontières (f), entre parenthèses, les combinaisons optimales des paramètres (ν, σ) pour chaque image



segmentations obtenues

distribution spatiale des erreurs (en noir)

a b c d

FIG. 2 – Résultats de SEROSVM sur les images tests (a - d) de la figure 1

5 Conclusion

La méthode, appelée SEROSVM, que nous avons proposée repose sur l'apprentissage d'un ensemble de *One-class SVM's* pour chaque classe. Chacun étant défini dans un sous-espace de représentation extrait aléatoirement de l'espace de représentation initial. Pour chacune des 2 classes et pour chaque espace de représentation, le *One-class SVM* permet d'estimer l'enveloppe d'une région contenant les observations considérées comme *normales* (donc appartenant à la classe). Une observation est hors de l'enveloppe d'une classe lorsqu'elle est issue de la classe adverse ou qu'elle est perturbée ou trop singulière. Dans ce cas, le *One-class SVM* l'étiquette comme *anormale* (vote nul) et n'intervient donc plus dans la décision finale (Cf. tableau 1 et formule 5).

L'efficacité de la méthode proposée a été étudiée dans le contexte de la segmentation d'images texturées. Nous avons montré que dans un environnement très *hétérogène* (avec des frontières très complexes), SEROSVM présente de très bonnes performances. Elle doit

permettre par ailleurs de s'adapter facilement aux situations plus complexes, notamment lorsque plus de deux classes sont en compétition. Cette généralisation est actuellement en cours d'expérimentation.

Bibliographie

Breiman, L. (2001). Random forests. *Machine Learning 45*(1), 5–32.

Brodatz, P. (1966). *Textures : A Photographic Album for Artists and Designers*. New York : Dover Publications.

Camci, F. and R. B. Chinnam (2008). General support vector representation machine for one-class classification of non-stationary classes. *Pattern Recognition 41*(10), 3021–3034.

Hayton, P., B. Schölkopf, L. Tarassenko, and P. Anuzis (2000). Support vector novelty detection applied to jet engine vibration spectra. In *Advances in Neural Information Processing Systems*, Volume 12, pp. 582–588.

He, X., P. Beauseroy, and A. Smolarz (2009). Nearest neighbour ensembles in lasso feature subspaces. *IET Computer Vision*. à paraître.

Kassab, R. and F. Alexandre (2009). Incremental data-driven learning of a novelty detection model for one-class classification with application to high-dimensional noisy data. *Machine Learning 74*(2), 191–234.

Li, X. and B. Liu (2003). Learning to classify texts using positive and unlabeled data. In *Proceedings of the Eighteenth International Joint Conference on Artificial Intelligence*, Acapulco, Mexico, pp. 587–594.

Manevitz, L. M. and M. Yousef (2001). One-class svms for document classification. *Journal of Machine learning research, Special Issue on Kernel Methods 2*, 139–154.

Schölkopf, B., J. C. Platt, J. Shawe-Taylor, A. J. Smola, and R. C. Williamson (2001). Estimating the support of a high-dimensional distribution. *Neural Computation 13*(7), 1443–1471.

Smolarz, A. (1997). Etude qualitative du modèle auto-binomial appliqué à la synthèse de texture. In *XXIXèmes journées de Statistique*, Carcassonne, pp. 712–715.

Yu, H., C. Zhai, and J. Han (2003). Text classification from positive and unlabeled documents. In *Proceedings of the twelfth International Conference on Information and Knowledge Management*, New Orleans, LA, USA, pp. 232–239.