

Apprentissage non supervisé des SVM par un algorithme des K-moyennes entropique pour la détection de zones brûlées

Olivier ZAMMIT, Xavier DESCOMBES, Josiane ZERUBIA

INRIA, projet commun ARIANA (INRIA-I3S)
2004, route des Lucioles, BP 93, 06902 Sophia Antipolis Cedex, France
Prenom.Nom@sophia.inria.fr

Résumé – L’objectif est d’évaluer les dégâts après un feu de forêt à partir d’une seule image satellitaire haute-résolution (SPOT5) acquise après le feu. Nous proposons une méthode de classification non supervisée fondée sur les SVM. La base d’apprentissage est définie automatiquement à partir d’une première classification obtenue par K-moyennes. L’algorithme K-moyennes utilisé pour l’apprentissage prend en compte un terme entropique qui permet de déterminer le nombre optimal de classes.

Abstract – In this paper, we assess the damaged caused by forest fire from a single after-fire and high resolution satellite image (SPOT5). An unsupervised method combining K-means and SVM is investigated to discriminate burnt from unburnt areas. In this technique, the training set is defined automatically by the K-means algorithm which takes into account an entropic term to determine the optimal number of classes.

1 Introduction

Les feux de forêts font partie des risques majeurs auxquels nous sommes régulièrement confrontés. En effet, chaque été en France, environ 30.000 hectares sont détruits. La détection des dégâts après un feu de forêt joue un rôle capital. Une évaluation précise des zones dévastées permet de planifier les programmes de restauration et de réhabilitation des sols, mais aussi de localiser les régions sinistrées, ce qui facilitera l’analyse du comportement du feu.

De nombreuses études ont montré l’efficacité des techniques de télédétection pour la détection des zones brûlées [1]. Elles sont généralement fondées sur l’analyse de changement d’indices de végétation, calculés à partir de plusieurs images satellitaires à faible-résolution, acquises avant et après le feu [2].

Le principal intérêt de notre méthode est d’estimer les zones brûlées à partir d’une seule image satellitaire. La classification entre zones brûlées et non brûlées est réalisée par Séparateurs à Vaste Marge (SVM) [3]. Cette technique de classification supervisée est particulièrement bien adaptée aux traitements de données de grandes dimensions et a démontré une grande robustesse et une bonne capacité de généralisation [4, 5]. De plus, elle a été récemment introduite dans les problèmes de classification d’images en télédétection [6, 7].

Cependant, les SVM nécessitent une phase d’apprentissage : cette technique consiste à trouver le séparateur optimal entre deux classes en utilisant un ensemble d’observations, appelé base d’apprentissage. La base d’apprentissage doit être changée de nombreuses fois afin d’optimiser expérimentalement les résultats.

Pour simplifier ce problème, on propose un processus de sélection automatique de la base d’apprentissage des SVM.

On effectue une première classification par une extension de l’algorithme des K-moyennes prenant en compte un terme d’entropie qui permet d’estimer le nombre de classes [8]. Cette première classification est utilisée pour définir automatiquement la base d’apprentissage.

Les résultats obtenus par association des K-moyennes et SVM sont comparés aux méthodes de classification plus traditionnelles automatiques (K-moyennes) et supervisées (K-plus proches voisins, SVM). Les zones brûlées extraites sont aussi comparées aux vérités de terrain fournies par le CNES et utilisées par le SDIS 83 (Service Départemental d’Incendie et de Secours du Var).

2 Description de la technique de classification utilisée

2.1 Algorithme des K-moyennes entropique

L’algorithme classique des K-moyennes propose de classer les données en K ensembles en minimisant l’énergie suivante :

$$U = \sum_{i=1}^K \sum_{x_j \in C_i} \|x_j - \mu_i\|^2$$

où μ_i désigne la moyenne des éléments de la classe C_i . Dans ce cas, le nombre de classes est fixé et représente le seul paramètre de l’algorithme, les moyennes μ_i étant estimées récursivement.

On définit l’entropie par $E = -\sum_{i=1}^K p_i \cdot \log(p_i)$, où $p_i = \frac{\text{Card}(C_i)}{N}$ représente la probabilité qu’un pixel de l’image de taille N appartienne à la classe C_i .

En ajoutant ce terme d'entropie à l'énergie précédente, on obtient :

$$U = \sum_{i=1}^K \sum_{x_j \in C_i} \|x_j - \mu_i\|^2 - \alpha_E \cdot \sum_{i=1}^K p_i \cdot \log(p_i)$$

On peut remarquer que le premier terme énergétique est minimal pour $K = N$, mais que le second terme est minimal lorsqu'il n'y a qu'une seule classe. Le paramètre α_E joue le rôle de compromis entre les deux termes. Après reformulation, on obtient :

$$U = \sum_{i=1}^K \sum_{x_j \in C_i} (\|x_j - \mu_i\|^2 - \alpha \cdot \log(p_i)) \text{ avec } \alpha_E = \alpha \cdot N$$

L'algorithme est initialisé avec un nombre élevé de classes. À chaque iteration, on assigne chaque pixel à la classe la plus proche au sens de la distance définie par l'équation précédente : $d(x_j, \mu_i) = \|x_j - \mu_i\|^2 - \alpha \cdot \log(p_i)$.

Il est à noter que le paramètre α décroît de manière exponentielle. Ainsi le terme entropique détermine dès les premières itérations le nombre optimal de classes mais n'intervient presque plus lorsque l'algorithme a convergé [9]. L'algorithme des K-moyennes entropique se comporte comme un simple algorithme des K-moyennes.

2.2 Sélection de la base d'apprentissage

L'algorithme des K-moyennes entropique est d'abord utilisé pour effectuer une première classification. Les différentes classes estimées sont alors regroupées en deux ensembles, l'une correspondant aux zones brûlées, l'autre aux zones non brûlées. Une opération d'érosion est réalisée afin d'éliminer les pixels mal classés et de ne garder qu'un petit nombre de pixels par classe. Ces pixels vont constituer la base d'apprentissage des SVM.

2.3 SVM

Cette technique fait partie des méthodes d'apprentissage supervisé et a été récemment introduite dans les problèmes de classification dans le domaine de la télédétection [7]. Elle consiste à trouver le séparateur optimal entre les différentes classes à partir d'un ensemble d'exemples appelé base d'apprentissage.

2.3.1 SVM linéaire

À partir d'un ensemble d'apprentissage (\vec{x}_i, y_i) où $\vec{x}_i \in \mathbb{R}^m$, ensemble des données et $y_i \in \{\pm 1\}$, celui des classes, les SVM consistent à trouver l'Hyperplan Séparateur Optimal (HSO) qui maximise la distance entre l'hyperplan et les deux classes [5]. Cette distance est appelée la marge.

L'hyperplan est défini par $\vec{w} \cdot \vec{x} + b = 0$ où (\vec{w}, b) désignent les paramètres de l'hyperplan (respectivement un vecteur normal au plan et le biais). Le classifieur est donné par : $f : \vec{x} \in \mathbb{R}^m \mapsto \text{sign}(\vec{w} \cdot \vec{x} + b) \in \{\pm 1\}$

Néanmoins, ce dernier doit satisfaire :

$$\begin{cases} \vec{w} \cdot \vec{x}_i + b \geq 0 & \text{si } y_i = +1 \\ \vec{w} \cdot \vec{x}_i + b \leq 0 & \text{si } y_i = -1 \end{cases}$$

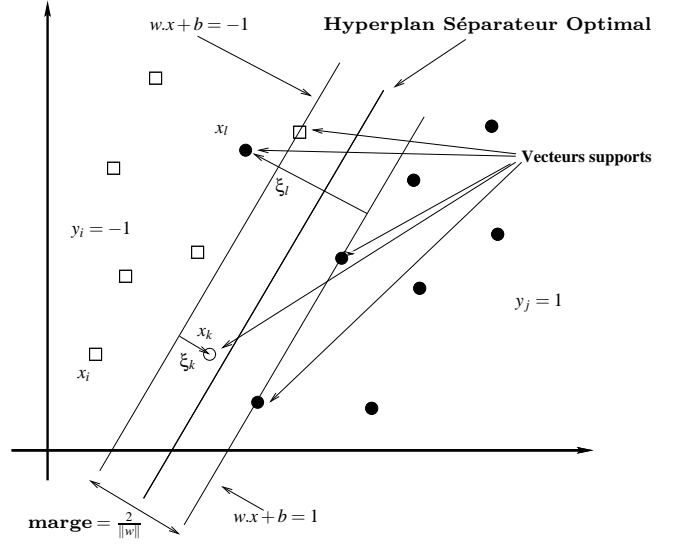


FIG. 1: classifieur SVM

De plus, le HSO doit maximiser la marge : $2/\|\vec{w}\|$. La solution est obtenue en résolvant le problème de maximisation de la marge sous les contraintes précédentes, ce qui mène au problème d'optimisation quadratique suivant :

$$\min_{(w,b)} \frac{\|\vec{w}\|^2}{2}$$

$$\text{tel que : } y_i(\vec{w} \cdot \vec{x}_i + b) \geq 1, \forall i \in \langle 1, N \rangle$$

En introduisant les multiplicateurs de Lagrange, on obtient le problème dual :

$$\max_{\lambda} W(\lambda) = -\frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \lambda_i \lambda_j y_i y_j \vec{x}_i \cdot \vec{x}_j + \sum_{i=1}^N \lambda_i$$

$$\text{tel que : } \sum_{i=1}^N \lambda_i y_i = 0, 0 \leq \lambda_i, \forall i \in \langle 1, N \rangle$$

qui est un problème d'optimisation quadratique convexe sous contraintes linéaires.

Le HSO est obtenu par $\vec{w} = \sum_{i=1}^N \lambda_i y_i \vec{x}_i$ et $b = y_i - \vec{w} \cdot \vec{x}_i$, avec i tel que $\lambda_i \neq 0$. La fonction de classification vaut donc :

$$f(\vec{x}) = \vec{w} \cdot \vec{x} + b = \sum_{i=1}^N \lambda_i y_i \vec{x}_i \cdot \vec{x} + b$$

Pour des données non linéairement séparables ou même dans le cas linéairement séparable, il peut être préférable de violer quelques contraintes pour agrandir la marge. Ainsi, on améliore la robustesse au détriment de l'erreur d'apprentissage.

Pour ce faire, on introduit les variables de relaxation ξ_i et le paramètre de régularisation C (voir Fig. 1). Le problème d'optimisation primal devient :

$$\min_{(w,b)} \frac{\|\vec{w}\|^2}{2} + C \sum_{i=1}^N \xi_i$$

$$\text{tel que : } y_i(\vec{w} \cdot \vec{x}_i + b) \geq 1 - \xi_i, \xi_i \geq 0, \forall i \in \langle 1, N \rangle$$

Ce problème d'optimisation est résolu à l'aide des multiplicateurs de Lagrange et devient :

$$\max_{\lambda} W(\lambda) = -\frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \lambda_i \lambda_j y_i y_j \vec{x}_i \cdot \vec{x}_j + \sum_{i=1}^N \lambda_i \quad (1)$$

tel que : $\sum_{i=1}^N \lambda_i y_i = 0$, $0 \leq \lambda_i \leq C$, $\forall i \in \langle 1, N \rangle$

2.3.2 SVM non linéaire

Dans la plupart des problèmes, l'hypothèse de linéarité est trop restrictive et le séparateur optimal doit pouvoir prendre une forme plus compliquée. La méthode du noyau est un moyen élégant et efficace pour traiter ce problème [10] : on transfère les données de l'ensemble de départ $\vec{x} \in \mathbb{R}^m$ vers un ensemble de dimension supérieure $\Phi(\vec{x}) \in \mathbb{R}^M$ dans lequel le problème devient séparable linéairement.

Le nouveau problème est obtenu en remplaçant les vecteurs de l'ensemble des données qui interviennent seulement dans le produit scalaire $\vec{x}_i \cdot \vec{x}_j$ par leurs correspondants dans le nouvel espace caractéristique $\Phi(\vec{x}_i) \cdot \Phi(\vec{x}_j)$. Ainsi, on introduit la fonction noyau dans l'équation (1) : $K(\vec{x}_i, \vec{x}_j) = \Phi(\vec{x}_i) \cdot \Phi(\vec{x}_j)$ et la nouvelle fonction de classification est donnée par :

$$f(\vec{x}) = \sum_{i=1}^N \lambda_i y_i K(\vec{x}_i, \vec{x}) + b$$

Il n'est cependant pas nécessaire de connaître explicitement la fonction Φ : toute fonction qui satisfait les conditions de Mercer [4] peut être utilisée comme noyau.

Dans les problèmes de classification d'images de télédétection, les noyaux les plus souvent utilisés sont :

- le noyau polynomial : $K(\vec{x}, \vec{x}') = (\vec{x} \cdot \vec{x}' + 1)^q$
- le noyau gaussien : $K(\vec{x}, \vec{x}') = \exp\left(-\frac{\|\vec{x} - \vec{x}'\|^2}{2\sigma^2}\right)$

Le point clé des SVM réside dans le choix du noyau et dans ses paramètres. Généralement, plusieurs noyaux et paramètres sont testés afin de sélectionner les meilleurs (en utilisant la technique de validation croisée par exemple).

3 Applications

Les données utilisées sont des images multispectrales SPOT 5 (Vert, Rouge et Proche Infra Rouge), de résolution 2.5m, acquises quelques jours après extinction des feux. Chaque image est le résultat de la fusion de l'image panchromatique Supermode de SPOT5 de résolution 2.5m et de l'image multispectrale de résolution 10m. Elles représentent la région du Massif des Maures (département du Var) qui a été gravement touchée par les feux lors de l'été 2003.

L'algorithme de SVM par apprentissage non supervisé par K-moyennes entropique est testé sur ces images et ses résultats sont comparés à ceux que fournissent les algorithmes de classification plus traditionnels (K-moyennes, K-plus proches voisins et SVM). Dans les deux SVM, on utilisera un noyau gaussien.

L'algorithme des K-Plus Proches Voisins (K-PPV) est une technique de classification supervisée très utilisée dans

TAB. 1: Performances des différents classifieurs

Vérité de terrain (2550 ha)	surface brûlée (ha)	précision (%)	FP (%)	FN (%)
K-moyennes	1967	97.89	0.86	2.21
K-PPV	1365	95.75	1.66	4.49
SVM	2284	98.73	2.15	1.19
Kmoyennes+SVM	2101	98.37	0.74	1.72

le domaine de la classification d'images. Elle assigne chaque pixel à la classe la plus représentée parmi les K plus proches éléments de la base d'apprentissage. Pour cet algorithme, la base d'apprentissage utilisée est la même que celle des SVM simple.

Pour chacune des quatre techniques de classification, comme les zones brûlées sont caractérisées par des ensembles homogènes et connexes, on régularise les classifications obtenues afin d'améliorer la cohérence spatiale des différentes zones et de réduire le nombre d'erreurs. Les images segmentées sont régularisées par un champ de Markov (modèle de Potts) [11].

Les zones brûlées obtenues par les quatre méthodes sont aussi comparées aux vérités de terrain fournies par un expert et obtenue par analyse manuelle de l'image satellitaire. Les résultats des différentes expérimentations sont listés dans le Tableau 1. Dans ce tableau, on y présente la surface brûlée totale, la précision de la classification, le taux de faux positifs et celui de faux négatifs.

Même si les différents algorithmes montrent un bon accord avec la vérité de terrain, les deux SVM offrent de meilleurs résultats et sont plus efficaces pour distinguer les zones brûlées des zones non brûlées. En effet, les K-moyennes et K-PPV sont moins sensibles aux objets qui présentent des caractéristiques spectrales très proches.

La classification par SVM avec apprentissage par K-moyennes permet d'avoir des résultats un peu plus précis que par simple K-moyennes, mais cependant moins précis que si la base d'apprentissage avait été choisie manuellement. La supériorité des résultats des SVM par rapport aux K-PPV démontre, par ailleurs, la capacité de bonne généralisation des SVM.

4 Conclusions

Dans cet article, nous nous sommes intéressés au problème de l'évaluation des dégâts après un feu de forêt à partir d'une seule image satellitaire haute résolution acquise après le feu. Une technique de classification automatique fondée sur l'algorithme des K-moyennes entropique et les SVM a été testée. Cette méthode utilise l'algorithme des K-moyennes entropique pour définir automatiquement la base d'apprentissage des SVM. Elle permet d'avoir des résultats plus précis que par simple K-moyennes.

Pour la suite, il est envisagé de tester l'influence des différents paramètres utilisés par les SVM mais aussi d'inclure l'information texturale de l'image afin d'améliorer les résultats de classification.

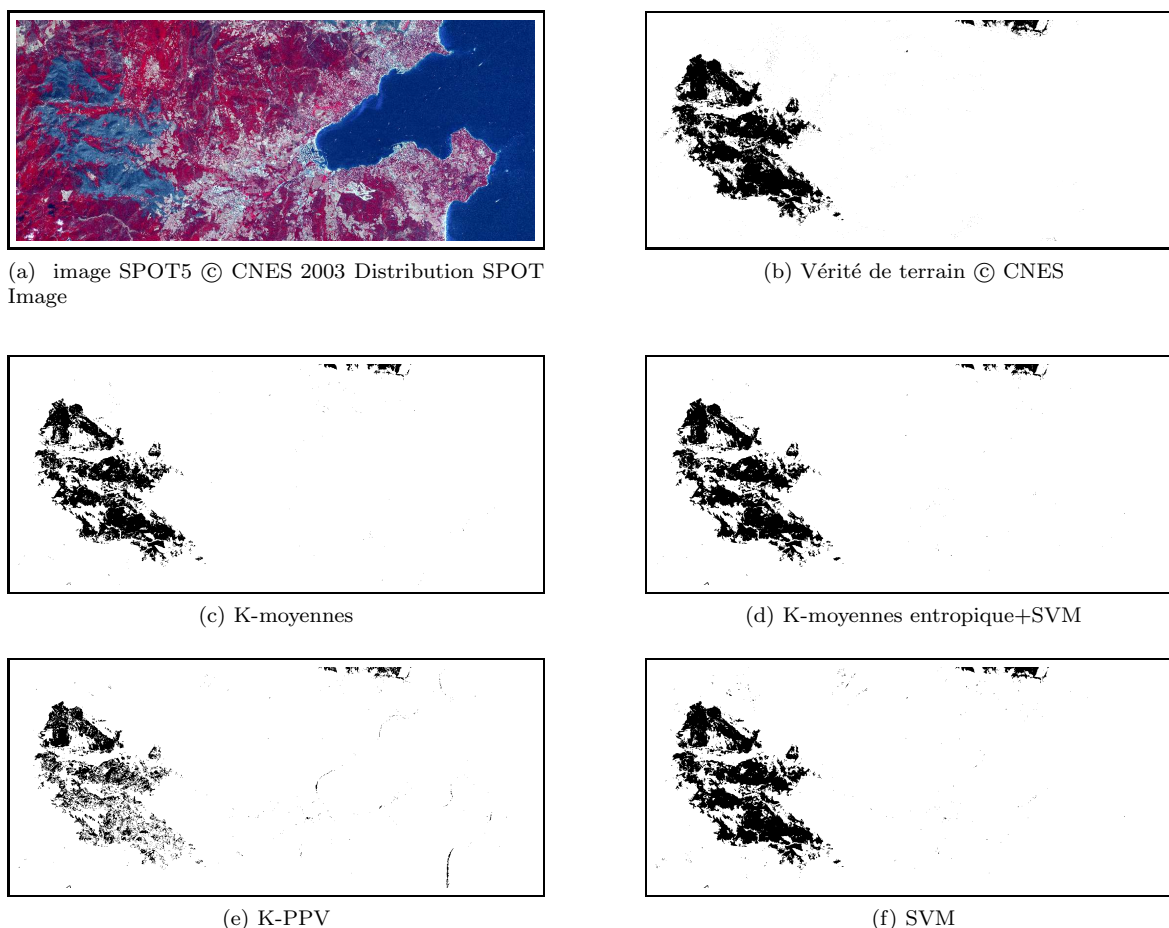


FIG. 2: Données et résultats expérimentaux obtenus par les différentes méthodes.

Remerciements

Les auteurs remercient la société Silogic pour le soutien financier partiel de ce travail ainsi que le CNES pour la fourniture de l'image satellitaire SPOT5 et de la vérité de terrain. Ils adressent aussi des remerciements au Commandant Jean-Claude Poppi, chef du service cartographie du SDIS du Var, pour de nombreuses discussions et des échanges fructueux.

Références

- [1] P.M. Barbosa, J. San-Miguel Ayanz, B. Martinez et G. Schmuck. *Burnt area mapping in southern Europe using IRS-WIFS. Forest Fire Research & Wildland Fire Safety*. Viegas (ed.). Millpress, Rotterdam, 2002.
- [2] E. Chuvieco, M.P. Martin et A. Palacios. *Assessment of different spectral indices in the red-near-infrared spectral domain for burned land discrimination*. International Journal of Remote Sensing, vol. 23, n° 23, 5103-5110, 2002.
- [3] O. Zammit, X. Descombes et J. Zerubia. *Burnt area mapping using support vector machines*. International Conference on Forest Fire Research, Figueira da Foz, Portugal, 2006.
- [4] V. Vapnik. *Statistical Learning Theory*. John Wiley and sons, inc., 1998.
- [5] B. Schölkopf, C.J.C. Burges et A.J. Smola. *Advances in Kernel Methods : Support Vector Learning*. MIT Press, 1999.
- [6] F. Melgani et L. Bruzzone. *Classification of hyperspectral remote sensing images with Support Vector Machine*. IEEE Trans. Geoscience and Remote Sensing, vol. 42, n° 8, 1778-1790, 2004.
- [7] F. Roli et G. Fumera. *Support Vector Machines for remote-sensing image classification*. Image and Signal Processing for Remote Sensing VI, SPIE Proceedings, vol. 4170, 160-166, 2001.
- [8] G. Palubinkas, X. Descombes et F. Kruggel. *An unsupervised clustering method using the entropy minimization*. IEEE International Conference on Pattern Recognition, Brisbane, Australie, 1998.
- [9] A. Lorette *Analyse de textures par méthodes markoviennes et par morphologie mathématiques : application à l'analyse des zones urbaines sur des images satellitaires*. Thèse de doctorat, Université de Sophia-Antipolis, septembre 1999.
- [10] B. Schölkopf, K. Tsuda et J.P. Vert, *Kernel Methods in computational biology*. MIT Press, 2004.
- [11] J.-P. Cocquerez et S. Philipp. *Analyse d'image : filtrage et segmentation*. Masson, Paris, 1995.