Publications du Groupe de travail STAPH: Statistique Fonctionnelle et Opératorielle

STAPH-2013-01 *Recueil de résumés de l'année 2012-2013*

Alain Boudou, Frédéric Ferraty, Yves Romain, Pascal Sarda, Philippe Vieu et Sylvie Viguier-Pla

Institut de Mathématiques, Université Paul Sabatier, Toulouse, France

.

Classification Supervisée en Imagerie Satellitaire Hyperspectrale

Mathieu Fauvel *

* Adresse pour correspondance:
Laboratoire DYNAFOR - UMR 1201
INRA & Université de Toulouse
e-mail: mathieu.fauvel@ensat.fr

Context and introduction

Recent advances in hyperspectral remote sensor technology allow the simultaneous acquisition of hundreds of spectral wavelengths for each image pixel. This detailed spectral information increases the possibility of more accurately discriminating and classifying materials of interest.

In the spectral domain, pixels are represented by vectors for which each component is a measurement corresponding to specific wavelengths [1]. The size of the vector is equal to the number of spectral bands that the sensor collects. For hyperspectral images, several hundreds of spectral bands of the same scene are typically available, while for multispectral images up to ten bands are usually provided. With increasing dimensionality of the images in the spectral domain, theoretical and practical problems arise. The idea of the *dimension* is intuitive, driven by experiments in one-, two- or three-dimensional spaces, and geometric concepts that are self-evident in these spaces do not necessarily apply in higher-dimensional spaces [2, 3]. For example, in high dimensional spaces, normally-distributed data have a tendency to concentrate in the tails, which seems to be contradictory with its bell-shaped density function [4]. Moreover, the rate of convergence of the statistical estimation decreases when the dimension grows while conjointly the number of parameters to estimate increases, making the estimation of the model parameters very difficult [5]. Consequently, with a limited training set, beyond a certain limit, the classification accuracy actually decreases as the number of features increases [6]. For the purpose of classification, these problems are related to the *curse of dimensionality*.

Several methods have been proposed in the literature to deal with hyperspectral data for the purpose of classification. A highly used strategy in remote sensing is *Dimension*

Reduction (DR). DR aims at reducing the dimensionality of data by mapping them onto another space of a lower dimension, without discarding any, or as less as possible, of the meaningful information. Recent overviews of DR can be found in [7, 8, 9]. Two main approaches can be defined.

1. Unsupervised DR: The algorithms are applied directly on the data without exploiting any prior information, and project the data into a lower dimensional space, according to some criterion (data variance maximization for PCA, independence for ICA ...).

2. Supervised DR: Training samples are available and are exploited to find a lower dimensional subspace where the class separability is improved. Fisher Discriminant Analysis (FDA) is surely one of the most famous supervised DR method.

It should be noted that supervised DR methods also face the "*curse of dimensionality*". For instance, FDA maximizes the ratio of the "between classes" scatter matrix, S_b, and the "within classes" scatter matrix, S_w. The optimal solution is given by the eigenvector corresponding to the first eigenvalues of $S_w^{-1}S_b$. In high dimensions, S_w^{-1} is in general ill-conditioned and, therefore, regularization is needed, which can be difficult to be tuned. Other popular DR methods such as Laplacian eigenmaps, Isomap or Locally Linear Embedding [8, Chapter 4 and 5] may be also limited by the dimensionality since they are based on the Euclidean distance between the samples. One last drawback of DR methods is the risk of losing relevant information. In general, DR methods act globally, which can be a problem for classification purpose: Different classes may be mapped onto the same subspace, even if the global discrimination criteria is maximized.

An alternative strategy to DR has been recently proposed: The subspace models [10]. These models assume that each class is located in a specific subspace and consider the original space without DR for the processing. For instance, the Probabilistic Principal Component Analysis (PPCA) [11] assumes that the classes are normally distributed in a lower dimensional subspace and are linearly embedded in the original subspace with additive white noise. Such models exploit the *empty space* property of hyperspectral data without discarding any dimension of the data [12, 13]. A general subspace model that encompasses several other models is the High Dimensional Discriminant Analysis (HDDA) model, proposed by Bouveyron *et al.* [14, 15].

Conversely, kernel based methods do not reduce the dimensionality but rather work with the full hyperspectral data [16]. These discriminative methods are known to be more robust to size of the dimensionality than conventional generative methods. However, local kernel methods are sensitive to the size of the dimensionality [17]. A kernel method is said to be local if the decision function value for a new sample depends on the neighbors of that sample in the training set. In high dimensional space the neighborhood of a sample is mostly empty, therefore such local methods are negatively impacted by the dimension. For instance, SVM

with Gaussian kernel

$$k_g(\mathbf{x}, \mathbf{z}) = \exp\left(- \frac{\|\mathbf{x} - \mathbf{z}\|^2}{2\sigma^2} \right)$$

is such a local kernel method.

Proposed approach

In this presentation, it is proposed to use subspace models to construct a kernel adapted to high dimensional data. The chosen approach for including subspace models in a kernel function is to consider the Mahalanobis distance, $d_{\mathbf{\Sigma}_c}$, between two samples for a given class, c, with covariance matrix, $\mathbf{\Sigma}_c$:

$$d_{\mathbf{\Sigma}_c}(\mathbf{x}, \mathbf{z}) = \sqrt{(\mathbf{x} - \mathbf{z})^t \mathbf{\Sigma}_c^{-1} (\mathbf{x} - \mathbf{z})}.$$

HDDA model is used for the definition of a class specific covariance matrix adapted for hyperspectral data. The specific signal and noise subspaces are estimated for each considered class, ensuring a parsimonious characterization of the classes. Following the HDDA model it is then possible to derive an explicit formulation of the inverse of the covariance matrix, without any regularization or dimension reduction. The parsimonious Mahalanobis kernel is constructed by substituting the Euclidean distance with the Mahalanobis distance computed using the HDDA model. Kernel hyperparameters are optimized using the so-called *radius margin* bound of the SVM classifier.

Outline

The presentation is organized as follows. In a first part, problems in hyperspectral image classification will be reviewed, then the proposed approach will be detailed in the second part. In the third part, experimental results on simulated and real data will be presented. Finally, perspectives on this work will be discussed. In particular, the CARNAGE (DYNAFOR-IMT) project on functional methods for the classification of hyperspectral image will be presented.

References

[1] C. Chang, *Hyperspectral imaging. Techniques for spectral detection and classification.* Kluwer Academic, 2003.

[2] D. Landgrebe, *Signal Theory Methods in Multispectral Remote Sensing.* New Jersey: John Wiley and Sons, 2003.

[3] M. Kendall, *A course in the geometry of n-dimensions.* New York: Dover Publication, 1961.

[4] L. Jimenez and D. Landgrebe, "Supervised classification in high-dimensional space: geometrical, statistical, and asymptotical properties of multivariate data," *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on*, vol. 28, pp. 39 –54, feb 1998.

[5] D. Donoho, "High-dimensional data analysis: the curses and blessing of dimensionality," in *AMS Mathematical challenges of the 21st century*, 2000.

[6] G. Hughes, "On the mean accuracy of statistical pattern recognizers," *IEEE Trans. Inf. Theory*, vol. IT-14, pp. 55–63, January 1968.

[7] C. J. C. Burges, "Dimension reduction: A guided tour," *Foundations and Trends in Machine Learning*, vol. 2, no. 4, pp. 275–365, 2010.

[8] J. A. Lee and M. Verleysen, *Nonlinear Dimensionality Reduction.* Springer, 2007.

[9] I. Guyon and A. Elisseeff, "An introduction to variable and feature selection," *J. Mach. Learn. Res.*, vol. 3, pp. 1157–1182, Mar. 2003.

[10] L. Parsons, E. Haque, and H. Liu, "Subspace clustering for high dimensional data: a review," *SIGKDD Explor. Newsl.*, vol. 6, pp. 90–105, June 2004.

[11] M. E. Tipping and C. M. Bishop, "Probabilistic principal component analysis," *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, vol. 61, no. 3, pp. 611–622, 1999.

[12] M. E. Tipping and C. M. Bishop, "Mixtures of probabilistic principal component analyzers," *Neural Computation*, vol. 11, no. 2, pp. 443–482, 1999.

[13] J. Baek, G. J. McLachlan, and L. K. Flack, "Mixtures of factor analyzers with common factor loadings: Applications to the clustering and visualization of high-dimensional data," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, pp. 1298 –1309, July 2010.

[14] C. Bouveyron, S. Girard, and C. Schmid, "High-Dimensional Discriminant Analysis," *Communication in Statistics- Theory and Methods / Communications in Statistics Theory and Methods*, vol. 36, pp. 2607 – 2623, Jan. 2007.

[15] C. Bouveyron, S. Girard, and C. Schmid, "High-Dimensional Data Clustering," *Computational Statistics and Data Analysis*, vol. 52, no. 1, pp. 502–519, 2007.

[16] T. Hofmann, B. Schölkopf, and A. J. Smola, "Kernel methods in machine learning.," *Annals of Statistics*, vol. 36, no. 3, pp. 1171–1220, 2008.

[17] Y. Bengio, O. Delalleau, and N. Le Roux, "The curse of highly variable functions for local kernel machines," in *Advances in Neural Information Processing Systems 18 (NIPS'05)* (Y. Weiss, B. Schölkopf, and J. Platt, eds.), pp. 107–114, MIT Press, 2006.

.

Les nouvelles théories de l'incertain

Didier DUBOIS

Adresse pour correspondance:
IRIT
Université Paul Sabatier
118, route de Narbonne
31062 Toulouse Cedex
e-mail: dubois@irit.fr

Résumé

Le but de cet exposé est de présenter les motivations qui ont poussé à chercher d'autres représentations de l'incertain, pour compléter celles utilisant des distributions de probabilité uniques. Il s'agit principalement de posséder des outils qui prennent fidèlement en compte l'information incomplète. On présentera ensuite quelques briques de base de la théorie des possibilités, de celles des fonctions de croyance et des probabilités imprécises ainsi que leurs liens. Puis on passera en revue les problèmes qu'on peut aborder avec ces théories. L'exposé est délibérément général pour situer les approches, au delà des détails techniques.

Mission d'Acquisition de DONNées hyperspectrales et lidar par capteurs Aéroportés

Anthony ZULLO *

* Adresse pour correspondance:
2 Impasse Paul Bressolles 31780 CASTELGINEST
e-mail: eric.zullo@orange.fr

Résumé

Le projet MADONNA (Mission d'Acquisition de DONNées hyperspectrales et lidar par capteurs Aéroportés) a pour but de construire un modèle afin de déterminer l'espèce à laquelle appartient un arbre lorsqu'on ne connaît que son hyperspectre.
L'objectif de la présentation concerne l'utilisation d'outils non-paramétriques fonctionnels afin d'essayer de résoudre ce problème à l'aide du logiciel R.
Un modèle de discrimination standard multilogit a d'abord été appliqué à ce jeu de données, avant de tester la discrimination non-paramétrique fonctionnelle.
Le critère de validation choisi est la minimisation de l'erreur de prédiction du modèle.
Un autre critère comparatif de la qualité du modèle est l'étude des spécificités et sensibilités de chacune des espèces.
Un autre outil de discrimination est actuellement en cours d'utilisation sur ces données : les SVM (Support Vector Machine).
D'autres critères seront également étudiés : la proportion de paires concordantes et les courbes ROC.

Références

F. Ferraty and P. Vieu, Nonparametric Functional Data Analysis: Theory and Practice. New York: Springer-Verlag, 2006.