Divergence de Kullback-Leibler en grande dimension pour la classification des prairies à partir de séries temporelles d'images satellite à haute résolution

Maïlys Lopes [1], Stéphane Girard [2] & Mathieu Fauvel [1]

[1] *DYNAFOR, Université de Toulouse, 31320 Auzeville-Tolosane, France - mailys.lopes@toulouse.inra.fr, mathieu.fauvel@ensat.fr*
[2] *Inria Grenoble Rhône-Alpes & LJK, 38334 Saint-Ismier, France - stephane.girard@inria.fr*

Résumé. Les nouvelles missions satellite offrent des séries temporelles d'images à haute résolution spatiale. Des outils statistiques appropriés sont requis afin de gérer la grande dimension des données face au faible nombre d'échantillons de référence. L'objectif de cette étude est de construire un modèle permettant la classification d'objets non-homogènes du paysage, les prairies, à partir d'une série temporelle d'un indice de végétation spectral. La méthode proposée utilise la divergence de Kullback-Leibler adaptée à la grande dimension pour calculer la distance entre chaque paire de prairies. Elle permettra la classification à l'échelle de l'objet avec un échantillon de petite taille et un nombre de variables élevé.

Mots-clés. Grande dimension, Divergence de Kullback-Leibler, Classification, Télédétection

Abstract. New satellite missions offer dense satellite image time series with high spatial resolution. Appropriate statistical tools are required to cope with the high dimension of the data and the small number of control samples. The aim of this study is to build a model that enables the classification of grasslands, non-homogeneous objects of the landscape, from a spectral vegetation index time series. The method uses the Kullback-Leibler divergence adapted to high dimension to compute the distance between each couple of grasslands. This model makes possible the classification at the object scale with a small sample size and a high number of variables.

Keywords. High dimension, Kullback-Leibler divergence, Classification, Remote sensing

1 Introduction

Les prairies sont un habitat semi-naturel d'intérêt pour les écologues car elles abritent une importante biodiversité. Les évènements de nature anthropiques, comme la fauche et/ou le pâturage perturbent le cycle naturel, la structure et la composition de la végétation. Pour ces raisons, il est important d'identifier les types de pratiques dans chaque parcelle afin de prédire leurs effets sur la biodiversité.

Dans ce contexte, la télédétection apparaît comme un outil adapté pour caractériser les prairies à l'échelle du paysage grâce à la large couverture spatiale et la fréquence de revisite élevée des capteurs satellite. Cependant, le signal réfléchi par les prairies est plus difficile à interpréter comparé aux couvertures végétales monospécifiques comme les cultures, car les prairies sont composées d'espèces diverses, variées et mélangées. Ce sont donc des objets non-homogènes. De plus, les prairies sont des éléments relativement petits du paysage (de l'ordre de l'hectare) qui nécessitent des images à haute résolution spatiale pour être détectables [1]. Etant donné leur cycle phénologique et la ponctualité des évènements anthropiques (*e.g.*, fauche), des séries temporelles très denses couvrant le cycle de végétation sont nécessaires pour identifier les types de conduite agricole [2].

Le but de cette étude est d'identifier les types de pratiques dans les prairies en utilisant des séries temporelles à haute résolution temporelle d'un indice de végétation spectral. La principale contrainte en télédétection est le faible nombre d'échantillons de référence comparé au grand nombre de variables.

Les approches classiques de classification opèrent pixel par pixel. Il en résulte plusieurs classes au sein d'une même parcelle, ce qui n'est pas réaliste, surtout dans le cas de modes de conduite qui sont définis à l'échelle de la prairie. Cela conduit à travailler à l'échelle de la parcelle, avec la contrainte qu'il ne s'agit pas d'un objet homogène dans le cas des prairies.

La première contribution de la méthode présentée est de prendre en compte la variabilité spectrale dans une prairie. Nous avons considéré que la distribution de l'indice de végétation des pixels dans une prairie donnée peut être modélisée par un distribution gaussienne. Puis, la divergence de Kullback-Leibler a été utilisée pour calculer la distance entre chaque paire de prairies. Pour faire face à la petite taille de l'échantillon de référence comparé au nombre de variables temporelles, un modèle parsimonieux gaussien est proposé en seconde contribution. Il s'agit d'une première étape pour faire de la classification supervisée et non supervisée.

2 Jeu de données

Le jeu de données est composé de 52 parcelles situées au sud de Toulouse, s'étendant sur 44000 km^2. Trois modes de conduite ont été identifiés sur cet échantillon : une fauche (34 parcelles), pâturage (10 parcelles) et mixte (fauche puis pâture, 8 parcelles). L'enquête a été effectuée en 2015 auprès des agriculteurs en les questionnant sur l'historique de la prairie.

Les données satellite sont issues d'une série de 17 images Formosat-2 acquises en 2013. Ce sont des données multispectrales (4 bandes) dotées d'une résolution spatiale de 8 mètres. Sur chaque image, nous avons calculé un indice de végétation, l'Indice de Végétation par Différence Normalisée (NDVI). Le NDVI reflète l'activité photosynthétique de la végétation. Ainsi, chaque pixel $\mathbf{x}$ de la prairie est representé par un vecteur de $\mathbb{R}^{17}$.

3 Divergence de Kullback-Leibler en grande dimension pour la classification

3.1 KLD symétrisée

La distribution de la réflectance des pixels des prairies est modélisée par une distribution gaussienne, *i.e.* la densité des pixels $\mathbf{x}$ est, conditionnellement à la prairie g_i, une distribution gaussienne. Pour calculer la similarité entre les distributions de deux prairies, nous avons utilisé la divergence de Kullback-Leibler (KLD) [3] symétrisée. La KLD symétrisée entre deux distributions gaussiennes peut s'écrire :

$$KLD(g_i, g_j) = \frac{1}{2}\Big[\operatorname{Tr}\left[\boldsymbol{\Sigma}_i^{-1}\boldsymbol{\Sigma}_j + \boldsymbol{\Sigma}_j^{-1}\boldsymbol{\Sigma}_i\right] + (\boldsymbol{\mu}_i - \boldsymbol{\mu}_j)^\top \big(\boldsymbol{\Sigma}_i^{-1} + \boldsymbol{\Sigma}_j^{-1}\big)(\boldsymbol{\mu}_i - \boldsymbol{\mu}_j)\Big] - d \quad (1)$$

où d est le nombre de variables, Tr est l'opérateur de la trace, $\boldsymbol{\mu}_i$ et $\boldsymbol{\Sigma}_i$ sont le vecteur moyen et la matrice de covariance de la prairie g_i. Ils sont estimés par leurs équivalents empiriques $\hat{\boldsymbol{\mu}}_i = \frac{1}{n_i}\sum_{l=1}^{n_i} \boldsymbol{x}_l$ et $\hat{\boldsymbol{\Sigma}}_i = \frac{1}{n_i}\sum_{l=1}^{n_i}(\boldsymbol{x}_l - \hat{\boldsymbol{\mu}}_i)(\boldsymbol{x}_l - \hat{\boldsymbol{\mu}}_i)^\top$ avec n_i le nombre de pixels dans la prairie g_i.

Malheureusement, le nombre de pixels utilisés dans l'estimation est petit comparé au nombre de variables. La Figure **??** montre que le nombre de pixels de la plupart des prairies est inférieur au nombre de paramètres à estimer. Ainsi, la matrice de covariance n'est pas inversible pour ces prairies. De plus, pour les autres prairies, les matrices de covariance estimées dans l'éq.(1) sont mal conditionnées, provoquant l'instabilité numérique de leur inverse. Pour faire face à ce problème, des développements statistiques spécifiques sont considérés dans le paragraphe suivant.

3.2 Divergence de Kullback-Leibler en grande dimension (HDKLD)

Dans ce travail, un modèle parcimonieux est utilisé pour modéliser la distribution gaussienne des prairies [4]. Ce modèle suppose que les dernières (les plus petites) valeurs propres de la matrice de covariance sont égales. Selon ce modèle, la matrice de covariance de la prairie g_i peut s'écrire :

$$\boldsymbol{\Sigma}_i = \mathbf{Q}_i \boldsymbol{\Lambda}_i \mathbf{Q}_i^\top + \lambda_i \mathbf{I}_d \quad (2)$$

où : $\mathbf{Q}_i = \left[\mathbf{q}_{i1}, \ldots, \mathbf{q}_{ip_i}\right]$, $\boldsymbol{\Lambda}_i = \operatorname{diag}\Big[\lambda_{i1} - \lambda_i, \ldots, \lambda_{ip_i} - \lambda_i\Big]$, $\mathbf{I}_d$ est la matrice identité de taille d, $\mathbf{q}_{ij}, \lambda_{ij}$ sont les j^{emes} valeurs/vecteurs propres de la matrice de covariance $\boldsymbol{\Sigma}_i$, $j \in \{1, \ldots, d\}$ tels que $\lambda_{i1} \geq \ldots \geq \lambda_{id}$, p_i est le nombre de valeurs propres non égales, et λ_i est la valeur propre multiple correspondant au bruit (dernières et égales valeurs propres).

Suivant ce modèle, l'inverse de la matrice de covariance peut être calculée explicitement :

$$\boldsymbol{\Sigma}_i^{-1} = -\mathbf{Q}_i \mathbf{V}_i \mathbf{Q}_i^\top + \lambda_i^{-1} \mathbf{I}_d \quad (3)$$

avec $\mathbf{V}_i = \text{diag}\Big[\frac{1}{\lambda_i} - \frac{1}{\lambda_{i1}}, \ldots, \frac{1}{\lambda_i} - \frac{1}{\lambda_{ip_i}}\Big]$, et l'éq.(1) peut s'écrire :

$$\text{HDKLD}(g_i, g_j) = \frac{1}{2}\Big[- \|\mathbf{\Lambda}_j^{\frac{1}{2}}\mathbf{Q}_j^\top\mathbf{Q}_i\mathbf{V}_i^{\frac{1}{2}}\|_F^2 - \|\mathbf{\Lambda}_i^{\frac{1}{2}}\mathbf{Q}_i^\top\mathbf{Q}_j\mathbf{V}_j^{\frac{1}{2}}\|_F^2$$
$$+\lambda_i^{-1}\,\text{Tr}\left[\mathbf{\Lambda}_j\right] - \lambda_j\,\text{Tr}\left[\mathbf{V}_i\right] + \lambda_j^{-1}\,\text{Tr}\left[\mathbf{\Lambda}_i\right] - \lambda_i\,\text{Tr}\left[\mathbf{V}_j\right] - \|\mathbf{V}_i^{\frac{1}{2}}\mathbf{Q}_i^\top(\boldsymbol{\mu}_i - \boldsymbol{\mu}_j)\|^2 - \|\mathbf{V}_j^{\frac{1}{2}}\mathbf{Q}_j^\top(\boldsymbol{\mu}_i - \boldsymbol{\mu}_j)\|^2$$
$$+ \frac{\lambda_i + \lambda_j}{\lambda_i\lambda_j}\|(\boldsymbol{\mu}_i - \boldsymbol{\mu}_j)\|^2 + \frac{\lambda_i^2 + \lambda_j^2}{\lambda_i\lambda_j}d\Big] - d \quad (4)$$

où $\|L\|_F^2 = \text{Tr}(L^\top L)$ est la norme de Frobenius.

3.3 Estimation

Les paramètres de l'éq.(4) sont estimés pour chaque prairie g_i à partir du vecteur moyen et la matrice de covariance empiriques tels que [4] :

- $\hat{\lambda}_i = \frac{\text{Tr}(\hat{\mathbf{\Sigma}}_i) - \sum_{j \leq p_i}\hat{\lambda}_{ij}}{d - p_i}$,
- $\hat{\lambda}_{ij}, \hat{\boldsymbol{q}}_{ij}$ sont les premiers valeurs/vecteurs propres de $\hat{\mathbf{\Sigma}}_i$, $j \in \{1, \ldots, p_i\}$. Ainsi, seul les p_i premiers valeurs/vecteurs propres sont nécessaires et l'estimation instable des vecteurs propres associés aux petites valeurs propres est évitée,
- $\hat{p}_i$ correspond au nombre de valeurs propres nécessaires pour atteindre un pourcentage de variance t donné, $\frac{\sum_{i=1}^{\hat{p}_i}\hat{\lambda}_i}{\sum_{i=1}^{d}\hat{\lambda}_i} \geq t$, t étant réglé durant l'apprentissage lors de la classification.

3.4 Utilisation de (HD)KLD en classification

La mesure (HD)KLD est une semi-métrique, *i.e.*, elle ne satisfait que les trois premiers axiomes d'une métrique [5] : (HD)KLD$(g_i, g_j) \geq 0$, (HD)KLD$(g_i, g_i) = 0$ et (HD)KLD$(g_i, g_j) =$ (HD)KLD(g_j, g_i). Cette semi-métrique peut être transformée en une fonction noyau définie positive en l'introduisant dans une fonction de base radiale [6] :

$K(g_i, g_j) = \exp\big[-\frac{(HD)KLD(g_i, g_j)^2}{\sigma}\big]$ avec $\sigma \in \mathbb{R}_+^*$. Ce noyau peut ensuite être utilisé en en tant que fonction noyau de SVM pour la classification supervisée.

4 Premiers résultats sur simulations

Afin de valider le modèle construit, nous avons tout d'abord procédé à une simulation numérique. Au cours de 100 répétitions, deux prairies g_i et g_j ont été simulées avec les paramètres suivants. La dimension d est égale à 17. Le nombre de pixels dans chaque prairie est tiré aléatoirement avec une probabilité égale parmi les déciles du nombre de pixels dans chaque prairie réelle (Figure **??**). p_i et p_j sont tirés aléatoirement suivant une loi uniforme discrète sur $\{1, 2, 3\}$. Les vecteurs moyens $\boldsymbol{\mu}_i$ et $\boldsymbol{\mu}_j$ sont simulés par des lois

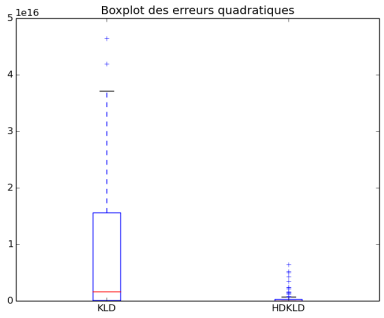


FIGURE 1 – Boxplot des erreurs quadratiques pour KLD et HDKLD. Les éléments aberrants de KLD ne sont pas affichés.

uniformes continues d-dimensionnelles standard, car le NDVI d'une prairie est compris entre 0 et 1. Les matrices de covariance $\mathbf{\Sigma}_i$ et $\mathbf{\Sigma}_j$ sont générées aléatoirement à partir des éq. (2) et (3) par la méthode suivante :

- Génération d'une matrice $\boldsymbol{M_i}$ aléatoire Gaussienne de taille (d, d),
- Décomposition $\boldsymbol{M_i} = \boldsymbol{Q_i R_i}$ avec $\boldsymbol{Q_i}$ une matrice orthonormale,
- Génération de $\mathbf{\Lambda_i}$ comme étant une matrice diagonale de taille (p_i, p_i) avec sur la diagonale un vecteur simulé par la loi p_i-dimensionnelle $\exp(-U[0, 3])$,
- λ_i est tiré aléatoirement suivant $\exp(-U[15, 20])$.

Pour chaque répétition, KLD (éq. 1) et HDKLD (éq. 4) sont calculés. Le boxplot des erreurs quadratiques des deux méthodes est montré sur la Figure 1. La racine carrée des erreurs quadratiques normalisées est RMSD = 2.32 pour KLD, et RMSD = 0.092 pour HDKLD. Ce résultat préliminaire montre l'instabilité de KLD face aux données de grande dimension. En revanche, HDKLD semble bien adaptée à ce type de données et donne une bonne estimation de la distance réelle.

5 Perspectives

La méthode sera appliquée sur le jeu de donnés réelles. Le noyau présenté au paragraphe 3.4 sera utilisé avec un SVM. Ainsi, les deux méthodes pourront être comparées au travers d'une classification supervisée sur le jeu de données.

Références

[1] Corbane, C., Lang, S., Pipkins, K., Alleaume, S., Deshayes, M., García Millán, V.E., Strasser, T., Vanden Borre, J., Toon, S. and Förster, M. (2015). Remote sensing for mapping natural habitats and their conservation status new opportunities and

challenges, *International Journal of Applied Earth Observation and Geoinformation*, vol. 37, pp. 7 - 16, Special Issue on Earth observation for habitat mapping and biodiversity monitoring.

[2] Schuster, C., Schmidt, T., Conrad, C.,Kleinschmit, B. and Förster, M. (2015). Grassland habitat mapping by intra-annual time series analysis comparison of RapidEye and TerraSAR-X satellite data, *International Journal of Applied Earth Observation and Geoinformation*, vol. 34, pp. 25 - 34.

Eilers, P.H.C. (2003). A perfect smoother, *Analytical Chemistry*, vol. 75, no. 14, pp. 3631-3636.

[3] Kullback, S. (1987). Letter to the editor : The Kullback-Leibler distance, *The American Statistician*, vol. 41, no. 4, pp. 340-341.

[4] Bouveyron, C., Girard, S. and Schmid, C. (2007). High-dimensional discriminant analysis, *Communications in Statistics - Theory and Methods*, vol. 36, no. 14, pp. 2607-2623.

[5] Aboumoustafa, K.T. and Ferrie, F.P. (2012). A note on metric properties of some divergence measures : The Gaussian case, *in Proc. of the 4th ACML, JMLR W&CP*, pp. 1-15.

[6] Haasdonk, B. and Bahlmann, C. (2004). Learning with distance substitution kernels, *in Pattern Recognition*, Rasmussen, C., Bülthoff, H., Schölkopf, B. and Giese, M. Eds., vol. 3175 of *Lecture Notes in Computer Science*, pp. 220-227.