Méthode conjointe de segmentation-classification pour des modèles d'écologie du déplacement

Marie-Pierre Etienne [1] , Julien Chiquet[1], Sophie Donnet [1] & Adeline Samson [2]

[1] *UMR MIA-Paris, AgroParisTech, INRA, Université Paris-Saclay, 75005, Paris, France, prénom.nom@agroparistech.fr*
[2] *LJK, Université Grenoble Alpes, France, adeline.leclercq-samson@univ-grenoble-alpes.fr*

Résumé.
En écologie du déplacement, une question classique consiste à identifier les activités d'un individu grâce à l'observation de sa trajectoire. En utilisant un modèle vectoriel auto-regressif comme modèle de mouvement, nous proposons une méthode de segmentation-classification en deux étapes, programmation dynamique puis régularisation, pour reconstruire l'activité au cours du temps.

Mots-clés. Trajectoires, segmentation, classification, méthodes régularisées, programmation dynamique

Abstract. A classical problem in movement ecology boils down to identify the activity of an animal with the observation of its movement. Using a vectoriel auto-regressive model, we propose a two steps segmentation clustering strategy, based on dynamic programming and regularization, for the reconstruction of the activity.

Keywords. Trajectories, segmentation, clustering, regularization , dynamic programming

1 Problématique

L'étude spatio-temporelle de la répartition des espèces est un champ majeur de l'écologie actuelle. Il peut être appréhendé au travers de l'étude des trajectoires, autrement nommée écologie du déplacement. La trajectoire témoigne entre autre des différentes activités d'un individu. Une des problématiques consiste alors à inférer le type d'activité de l'animal à partir de telles observations. Ce problème peut également apparaître dans un contexte halieutique.

On appelle trajectoire une suite de triplets (t, Y_t^1, Y_t^2) où t est le temps, Y_t^1 et Y_t^2 sont la longitude et latitude respectivement. Des modèles à temps continu et à temps discret ont été utilisés pour décrire ce type de données (voir [1]). Le plus souvent le lien entre activité et trajectoire est introduit par un modèle à espace d'états (par exemple modèles de chaînes de Markov cachées dans les cas à temps discret, [2]). Une des limites de ce type d'approche est l'hypothèse de temps exponentiel passé dans chacune des activités.

Dans ce travail nous nous intéressons aux modèles à temps discrets et proposons une méthode de détection de rupture (détection de changement d'activité) par segmentation de façon à s'abstraire d'une hypothèse sur la longueur du temps passé dans chaque activité. Cette méthode est adaptée au cadre bivarié spécifique des données de trajectoires. À l'issue de l'étape de segmentation, un segment est associé à une pseudo activité. Nous proposons d'utiliser les techniques de régularisation pour regrouper les segments correspondant à une même signature d'activité.

2 Méthode

Notons $Y_t = (Y_t^1, Y_t^2)$. Soient $(Y_0, \ldots, Y_n) \in (\mathbb{R}^2)^{n+1}$ les positions aux instants $t_0, t_1, \ldots, t_n$ $(t_0 = 0)$ tels que pour tout $j \in [1, n]$, $t_i - t_{i-1} = \Delta$. Parmi les modèles de mouvement en écologie, le modèle vectoriel auto-régressif (VAR) s'écrit de la façon suivante :

$$Y_{i+1} = RY_i + \nu + E_{i+1}, \quad E_i \overset{i.i.d}{\sim} \mathcal{N}(0, \Psi),$$

où $\nu \in \mathbb{R}^2$ et Ψ est une matrice de covariance de taille 2×2 et $R \in \mathcal{M}_{2,2}$ est une matrice dont la norme spectrale est strictement inférieure à 1.

L'extension de ce modèle à plusieurs activités prend la forme suivante. Soient A le nombre d'activités et $J \geq A$ le nombre de changement d'activités. On note $\tau_0 = 0 < \tau_1 < \cdots < \tau_j < \cdots < \tau_J = n$ les instants de changement d'activité et $a_j \in \{1 \ldots A\}$ l'indice de l'activité sur le segment $S_j = \{\tau_{j-1} + 1, \ldots, \tau_j\}$. Alors pour tout $i \in S_j$, on écrit :

$$Y_{i+1} = R_{a_j} Y_i + \nu_{a_j} + E_{i+1}, \quad E_i \overset{i.i.d}{\sim} \mathcal{N}(0, \Psi_{a_j}).$$

On s'intéresse à l'estimation conjointe de la partition $(S_j)_{j=1,\ldots,J}$, et des paramètres correspondant à chaque activité $(\nu_a, R_a, \Psi_a)_{a=1\ldots A}$. Nous proposons une méthode en deux étapes : segmentation puis régularisation.

- **Segmentation** Dans un premier temps, nous cherchons une segmentation définie par la partition $(S'_k)_{k=1,\ldots K}$ de $\{1, \ldots, n\}$, $K \geq J$, telle que :

  $$\forall i \in S'_k, \quad Y_{i+1} = R_k Y_i + \nu_k + E_{i+1}, \quad E_i \overset{i.i.d}{\sim} \mathcal{N}(0, \Psi_k),$$

  À cette étape, à chaque segment S'_k est associé à un jeu de paramètres (ν_k, R_k, Ψ_k). La partition $(\hat{S}'_k)$ et les paramètres $(\hat{\nu}_k, \hat{R}_k, \hat{\Psi}_k)_{k=1\ldots K}$ maximisant la vraisemblance s'obtiennent par un algorithme de programmation dynamique ([3]).

- **Régularisation** A partir de cette sur-segmentation, l'étape de régularisation a pour but de regrouper les segments similaires, signatures d'une activité. Afin de pourvoir utiliser les méthodes de régularisation développées dans un cadre d'observations indépendantes, une première étape consiste à blanchir le signal au sein de chaque segment obtenu. Nous définissons donc:

$$\forall i \in \hat{S}'_k, \quad \tilde{Y}_i = \hat{\Psi}_k^{-1/2} Y_i$$

Sur le signal $(\tilde{Y}_i)_{i=1\ldots n}$ bivarié et blanchi, nous proposons une pénalité type fused lasso multivariée permettant d'attribuer à chaque segment une des A activités [4].

La pertinence de la méthode est illustrée sur données simulées et réelles.

Bibliographie

[1] McClintock, B. T., Johnson, D. S., Hooten, M. B., Ver Hoef, J. M., et Morales, J. M. (2014). *When to be discrete: the importance of time formulation in understanding animal movement. Movement ecology*, 2(1), 21.
[2] Patterson, T. A., Thomas, L., Wilcox, C., Ovaskainen, O., et Matthiopoulos, J. (2008). *State–space models of individual animal movement.* Trends in ecology & evolution, 23(2), 87-94
[3] Cleynen, A. (2013). *Statistical Approaches for Segmentation: Application to Genome Annotation* (Doctoral dissertation, Université Paris Sud-Paris XI).
[4] Hocking, T., Vert, J.P., Joulin, A., Bach F.R. (2011). *Clusterpath: an Algorithm for Clustering using Convex Fusion Penalties* . Proceedings of the 28th International Conference on Machine Learning (ICML-11),745–752