

Prédiction du niveau de pauvreté par Deep Learning à partir d'une séquence d'image satellites

Sujet niveau Master 2, à partir de février 2019

**Marc Chaumont, Gérard Subsol
Laure Berti-Équille et Dino Ienco**

LIRMM (Laboratoire d'Informatique, de Robotique et Microélectronique de Montpellier) Equipe ICAR
Bât. 5, CC 05016 - 860 rue de St Priest
34095 Montpellier cedex 5, France
Tel : +33 4.67.14.97.59
Marc.Chaumont@lirmm.fr

Mots clefs : Imagerie satellite, Deep-Learning, série temporelle, multi-résolution, classification, données manquantes, traitement d'image.

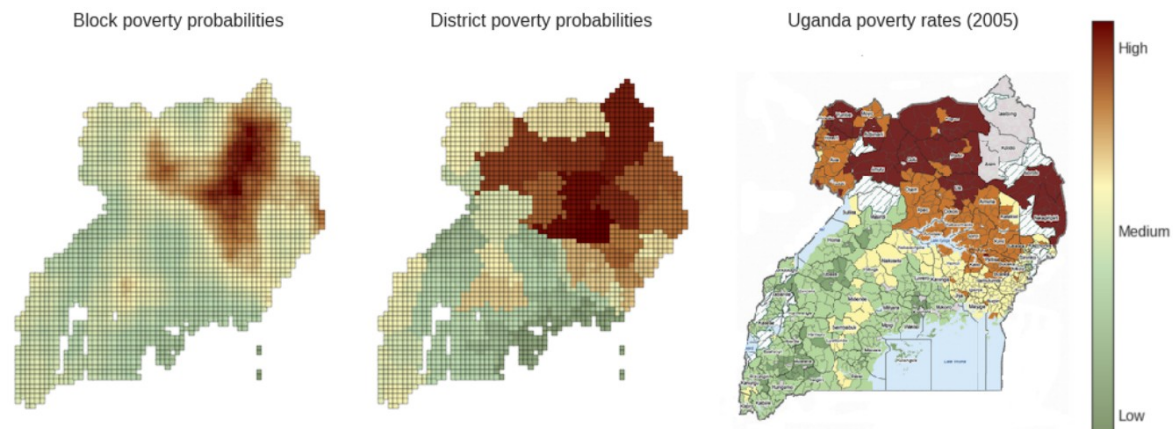


Figure 3: **Left:** Predicted poverty probabilities at a fine-grained 10km × 10km block level. **Middle:** Predicted poverty probabilities aggregated at the district-level. **Right:** 2005 survey results for comparison (World Resources Institute 2009).

Figure extraite de [Xie et al 2016 - Transfer]

Bien souvent en *télédétection*, on souhaite produire une cartographie géographique, c'est-à-dire produire une carte découpée en zones, où chaque zone est *labellisée* par une valeur. Cette valeur peut par exemple être le niveau de consommation annuel des ménages, la santé des actifs, etc.

L'approche standard pour produire une cartographie consiste à extraire des caractéristiques / traits / informations pour chaque zone étudiée. Ces caractéristiques / traits / informations sont représentés par un ensemble de valeurs entières, réelles ou binaires. Ces caractéristiques peuvent être le matériau des bâtiments, le matériau des toits, le nombre de pièces d'une maison, le type de maison, les distances entre divers points d'infrastructure, la classification urbaine ou rurale, la température annuelle, les précipitations annuelles, etc. [Xie et al 2016 - Transfer]. Il est bien évident qu'il n'est pas aisé d'avoir accès à de telles informations qui nécessitent de faire des relevés ou des enquêtes sur le terrain.

Une solution moins coûteuse pour faire la cartographie consiste à utiliser des *images satellites de haute résolution* (le principe de la *télédétection* consiste à faire de la mesure à distance), et d'extraire des caractéristiques qui pourront alors servir à prédire la valeur de chaque zone de la carte. Par exemple, en 2016, Xie et al. [Xie et al 2016 - Transfer] ont proposé de prédire le niveau de pauvreté (= niveau de consommation annuel des ménages) pour des zones de 1 km x 1 km (voir la figure ci-dessus). Toute la problématique de la cartographie par image satellite est qu'il est nécessaire d'avoir suffisamment d'images labellisées (images + valeurs pour chaque zone) pour pouvoir utiliser des algorithmes d'apprentissage automatique. La publication de Xie et al. a cela d'intéressant qu'elle ne nécessite que très peu d'images labellisées puisqu'elle repose sur une approche en deux étapes :

1. un apprentissage préalable d'un CNN (*Convolutional Neural Network*) pour prédire l'intensité lumineuse dans des images satellite prises de nuit, et cela, à partir d'images satellite prises de jour,
2. puis une réutilisation de ce réseau CNN (notion de "*Transfer Learning*") pour cette fois-ci lui apprendre à prédire la pauvreté (utilisation d'une régression) à partir d'image satellite de jour. En effet, on va supposer que le niveau de pauvreté est inversement corrélé à l'illumination nocturne (plus il y a d'électricité, moins la zone est pauvre).

L'approche de Xie et al. est particulièrement élégante puisqu'elle ne nécessite plus d'interventions sur le terrain pour obtenir des caractéristiques. Elle permet également de faire une prédiction sur des pays entiers comme par exemple l'Afrique (voir les résultats dans [Jean et al. 2016 - PredictPoverty]). Sur leur ensemble de tests, l'approche de Xi et al. obtient 71% de précision, ce qui est meilleur de 3% par rapport aux approches sans transfert, utilisant également du Deep-Learning et elle est seulement 4% inférieure à l'approche reposant sur une utilisation d'un relevé sur le terrain. Dans une publication plus récente, Jean et ses collaborateurs [Jean et al. 2019 - Tile2Vec], toujours dans un esprit d'approche faiblement supervisé, proposent un apprentissage par "triplet loss". Cela correspond à apprendre avec des triplets d'images (deux zones proches dans l'image satellite et une zone distante), un vecteur caractéristique discriminant la zone proche de la zone distante. Cet apprentissage est ainsi totalement non supervisé. Dans un deuxième temps, une régression est effectuée pour prédire le niveau de pauvreté à partir du vecteur des caractéristiques de l'image en entrée. Les résultats de cette nouvelle proposition donnent une régression dont la corrélation est meilleure que celle obtenue par l'approche par transfert [Xie et al 2016 - Transfer]. Il reste cependant encore une marge de progression puisque la corrélation après régression n'est que de 70%.

La méthodologie est intéressante, mais comme indiqué par les auteurs [Jean et al. 2019 - Tile2Vec], elle ne prend pas assez en compte l'aspect temporel. Par ailleurs, les données sur lesquelles les expériences ont été effectuées ont été volontairement bruitées par les agences gouvernementales.

L'étudiant devra donc étudier et proposer une solution dans le cas où l'on dispose d'un ensemble d'images satellites prises sur une dizaine d'années, avec une fréquence variable, un échantillonnage à "trous", une échelle en résolution variable et un faible nombre de données annotées.

Dans un premier temps, l'étudiant devra faire un état de l'art des approches de segmentation (labellisation) d'images satellites, en veillant à étudier plus particulièrement les approches de prédiction utilisant des séquences d'images satellite. En parallèle, l'étudiant reprendra les expériences menées par l'équipe de Jean et al. pour avoir une base de comparaison, mais également pour prendre en main les données dont nous disposons. En fin de stage, l'étudiant pourra comparer l'approche de Jean et al. et son approche basée sur la prédiction par séquence d'images.

Pré requis : programmation C/C++, Python, Classification, Fouille de données, connaissances en traitement d'images, connaissances de base en Deep Learning, anglais écrit scientifique. Aucune connaissance en télédétection n'est requise.

Conditions de stage :

Durée : 5-6 mois

Indemnités : ~600 € / mois. Financé par le CESAB sur l'appel Belmont.

Le stage se déroulera au LIRMM (campus St Priest) dans l'équipe ICAR.

Références :

- <https://nealjean.com/>

- [Jean et al. 2019 - Tile2Vec] [Tile2Vec: Unsupervised representation learning for spatially distributed data](#) N. Jean, S. Wang, A. Samar, G. Azzari, D. Lobell, S. Ermon *AAAI Conference on Artificial Intelligence (AAAI)*, 2019

- [Jean et al. 2016 - PredictPoverty] [Combining satellite imagery and machine learning to predict poverty](#) N. Jean, M. Burke, M. Xie, W. M. Davis, D. B. Lobell, S. Ermon, *Science*, 353(6301), 790-794, 2016. Voir site web et supplementary material

https://static1.squarespace.com/static/57a8ec72c534a5448c606796/t/57b74438893fc005181274ad/1471628351213/JeanBurkeEtAl2016_SI.pdf

- [Xie et al 2016 - Transfer] [Transfer learning from deep features for remote sensing and poverty mapping](#) M. Xie, N. Jean, M. Burke, D. B. Lobell, S. Ermon *AAAI Conference on Artificial Intelligence (AAAI)*, 2016.