

Nom : Thomas Guyet¹, René Quiniou²

Équipe :

1. Laboratoire d'Informatique AGROCAMPUS-OUEST – IRISA Équipe DREAM
2. INRIA – IRISA Équipe DREAM

Mail : thomas.guyet@agrocampus-ouest.fr, quiniou@irisa.fr

Lien : http://www.irisa.fr/dream/Pages_Prof/Thomas.Guyet/Stages/

Contacts : Thomas Guyet, René Quiniou

Titre : Réalisation d'une extension RapidMiner pour la fouille de données sur des images satellite

Mots clés : intelligence artificielle, fouille de données, traitement de données géographiques, images satellite, application agronomique.

Description :

Les images satellite permettent l'acquisition d'information à grande échelle sur les couvertures de sol. Elles se différencient des images classiques (trois couches RGB), d'une part, parce que chaque pixel d'une telle image peut contenir plus de couches et, d'autre part, parce que les pixels sont géoréférencés.

Nous nous intéressons en particulier aux séries temporelles d'images satellite (SITS), c'est-à-dire des images composés d'une « couche » acquise tous les seize jours sur plusieurs années. Les travaux récents de l'équipe DREAM [1] ont permis de mettre en place une méthode pour segmenter une SITS en utilisant des méthodes de fouille de données (clustering, fouille de séquence, ...). Dans ce premier travail, nous avons expérimenté un algorithme de clustering, alors que d'autres algorithmes pourraient donner de meilleurs résultats. Nous sommes donc intéressé par expérimenter une multitude d'algorithme et par en évaluer les résultats.

Pour tester différents algorithmes de clustering, nous ne souhaitons pas réimplémenter tous ces algorithmes, mais plutôt utiliser les implémentations existantes dans des boîtes à outils tels que RapidMiner [2]. Ce logiciel permet de définir une « chaîne de traitement » pour réaliser une fouille de données ainsi que des plans d'expérimentation pour comparer les performances des méthodes. Le logiciel propose une interface pour connecter des opérateurs (algorithmes de fouille de données ou de transformation de données).

Dans ce stage, nous souhaitons réaliser une « extension » RapidMiner (ou plugins), c'est-à-dire un ensemble d'opérateurs qui peuvent être utilisés dans l'interface de RapidMiner, pour intégrer le traitement d'images géoréférencées.

Dans un premier temps, le stagiaire aura pour objectif de définir les opérateurs nécessaires à la construction de « chaînes de traitement » d'images géoréférencées et de mettre en place les opérateurs principaux :

- opérateurs de chargement d'images dans un format géoréférencé (e.g. GeoTiff) et de chargement d'images au format vectoriel (shapefile),
- opérateur efficace pour la mise en relation de plusieurs images,
- opérateurs d'enregistrement réaliser des opérations de classification des pixels d'images satellite).

Une fois ces opérateurs mis en place, ils pourront être mis en relation avec les autres opérateurs de RapidMiner de sorte à construire une chaîne de traitement pour la classification de SITS. La réalisation pourra être testée sur des données issues de séries temporelles d'images satellite (données du Sénégal).

Les développements seront réalisés en Java. Une expérience dans ce langage est souhaitable. Les technologies utilisées vont permettre à l'étudiant de découvrir, d'une part, les méthodes de fouille de données utilisées dans RapidMiner, les méthodes spécifiques de traitement de

l'information géographique (utilisation de la librairie GeoTools).

Le stage sera effectué conjointement au Laboratoire d'Informatique d'Agrocampus-Ouest et à l'IRISA au sein de l'équipe DREAM (Diagnostic, Recommandation d'Action, Modélisation).

Références :

- [1] Guyet T., Nicolas H., Diouck A. « Segmentation multi-échelle de séries temporelles d'images satellite : Application à l'étud/ d'une période de sécheresse au Sénégal », RFIA, 2011.
- [2] RapidMiner : www.rapidminer.com