

Sujet de Master 2 Recherche ou équivalent École d'ingénieurs

Apprentissage collaboratif multi-paradigme pour l'analyse de séquences d'images de télédétection

<i>Rattachement :</i>	AgroParisTech - Paris
<i>Direction :</i>	Antoine CORNUÉJOLS (AgroParisTech) (antoine.cornuejols@agroparistech.fr) Pierre GANÇARSKI (ICube, Université de Strasbourg) (gancarski@unistra.fr) Emmanuelle SARRAZIN (CNES) (Emmanuelle.Sarrazin@cnes.fr)
<i>Lieu du stage :</i> (au choix)	UMR MIA (Mathématiques et Informatique Appliquées) 16, rue Claude Bernard, 75005 Paris
	ou
	UMR ICUBE - Université de Strasbourg 300 bld Sébastien Brant, 6412 Illkirch
<i>Durée</i>	5 à 6 mois (entre février et juillet 2021)

1. Sujet

Ce stage de Master 2, très fortement orienté Recherche, s'inscrit dans le cadre d'un projet co-porté par les laboratoires MIA (AgroParisTech) et ICube (Université de Strasbourg) en collaboration étroite avec le CNES (Centre National d'Études Spatiales). Ce projet a pour but de développer des méthodes innovantes en apprentissage automatique pour l'étude et l'interprétation de séries d'images de télédétection en vue d'extraire et mettre en évidence les phénomènes sous-jacents affectant les zones étudiées (urbanisation, occupation des sols, pratiques agricoles, érosion...). Il s'agit d'un problème difficile pour lequel des algorithmes peuvent s'appuyer sur des données variées, dont certaines, assez peu nombreuses, ont été étiquetées par des experts, et d'autres, beaucoup plus nombreuses, ne l'ont pas été.

L'objectif de ce stage est d'étudier les concepts entrant dans la définition de méthodes d'apprentissage collaboratif multi-paradigme pour des données temporelles dans lesquelles des agents d'apprentissage supervisé et des agents d'apprentissage non supervisé collaborent afin d'améliorer mutuellement leur résultat et, in fine, de parvenir à une interprétation consensuelle de ces données.

Il s'agira en particulier de répondre des questionnements scientifiques tels que :

- Quelles informations (données, hypothèses, contraintes...) les méthodes doivent elles échanger pour mettre en place cette collaboration ?
- Comment évaluer globalement ou deux à deux des résultats par nature différent (modèle vs regroupement par exemple) ?
- Comment en déduire globalement ou au niveau des agents, les actions à entreprendre pour améliorer les résultats et/ou consensus courants ?
- Comment combiner les avis des différents agents en un consensus ?
- Dans un scénario d'échanges itératifs, comment assurer une convergence ?

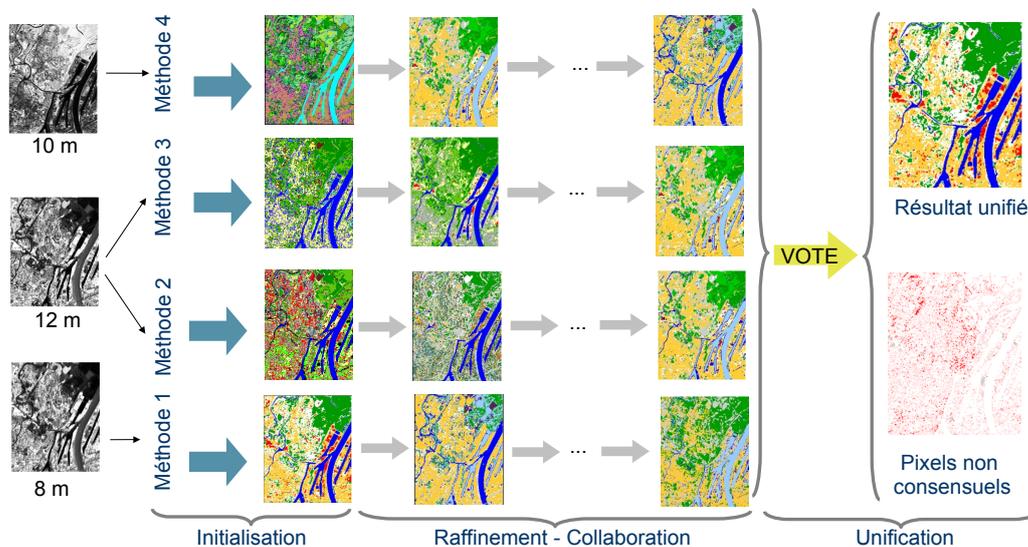


FIGURE 1 – Clustering collaboratif

La littérature propose déjà un certain nombre de réponses plus ou moins partielles pour des apprentissages particuliers sur lesquelles le travail de stage pourra s'appuyer. En apprentissage supervisé, on citera les méthodes d'ensemble (e.g. boosting [SF12], co-learning [BM98] ou les approches semi-supervisées [CSZ10]. Par contre, hormis la méthode SAMARAH [FGW10] (Fig. 1) proposée par ICube, très peu de résultats sont établis dans le cadre de méthodes essentiellement non supervisées, en particulier dans le cas du clustering collaboratif (voir [CWGB18] pour un panorama de questions et d'approches). Mais, ces cadres sont tous très contraints et ne s'étendent pas directement à l'étude de l'apprentissage collaboratif multi-paradigme avec des données hétérogènes mesurées dans le temps.

Les **résultats attendus** sont les suivants :

- Un *état de l'art* synthétique sur les méthodes d'apprentissage collaboratif multi paradigme avec mise en perspective par rapport aux méthodes collaboratives en général.
- La *définition de protocoles d'apprentissage collaboratif* entre une méthode d'apprentissage supervisé et une méthode d'apprentissage non supervisé collaborent dans le cadre de l'analyse de séries temporelles d'images de télédétection.
- Une expérimentation de certains de ces protocoles sur des séquences d'images de télédétection. Les zones étudiées seront définies en accord avec les experts du CNES. On pourra s'appuyer sur la plateforme FoDoMuST¹ actuellement développée par ICube et/ou sur l'Orféo Tool Box proposée par le CNES.
- Si possible, la détermination de *garanties* théoriques sur les propriétés de certains de ces protocoles.

Le stagiaire travaillera en collaboration et avec le support des chercheurs et des autres stagiaires

1. <https://sdc.icube.unistra.fr/en/index.php?title=FODOMUST>

participant à ce projet. Il rencontrera des spécialistes de la télédétection ainsi que des experts des domaines d'application (Géographie, Géosciences, Hydrologie...) et se verra confronté à des problématiques réelles en Sciences de l'Environnement.

Ce projet dispose d'un co-financement du CNES et du projet ANR HERELLES² pour une thèse débutant en octobre 2021. Ce stage se situe en amont de cette thèse et constitue un tremplin vers celle-ci.

2. Pré-requis

- Avoir de solides connaissances en apprentissage automatique.
- Avoir de bonnes notions d'analyse de séries temporelles
- Être à l'aise dans l'écriture de programmes Python et/ou en JAVA
- Avoir une appétence forte pour la recherche en informatique et dans le domaine de la télédétection
- Être curieux et autonome, tout en ayant le goût du travail en équipe

Une connaissance en analyse d'images de télédétection est un plus mais n'est pas obligatoire, une formation à l'analyse de ce type de données sera assurée.

3. Environnement du stage

Le stage se déroulera en collaboration forte entre les deux laboratoires et le CNES. Le stagiaire pourra choisir son lieu de stage (MIA ou ICube).

Ce stage sera gratifié sur la base légale (≈ 600 €/mois).

4. Pour postuler

Envoyez un mail à antoine.cornuejols@agroparistech.fr et (impérativement) pierre.gancarski@unistra.fr en joignant :

1. une *lettre de motivation* expliquant vos qualifications, expériences et motivations ;
2. un *curriculum vitae* ;
3. toutes les informations sur votre cursus universitaire : parcours, diplômes obtenus, relevés de notes ainsi que votre *classement* pour chacune des années de Master ou équivalent pour les écoles d'ingénieurs ;
4. si possible, les coordonnées des personnes (enseignants ou autres personnes) pouvant fournir des informations sur vos compétences, votre motivation et votre travail.

2. herelles-anr-project.cnrs.fr

Références

- [BM98] Avrim Blum and Tom Mitchell. Combining labeled and unlabeled data with co-training. In *Proceedings of the eleventh annual conference on Computational learning theory*, pages 92–100, 1998.
- [CSZ10] Olivier Chapelle, Bernhard Schölkopf, and Alexander Zien, editors. *Semi-supervised Learning*. MIT Press, Cambridge, MA, USA., 2010.
- [CWGB18] Antoine Cornuéjols, Cédric Wemmert, Pierre Gançarski, and Younès Bennani. Collaborative clustering : Why, when, what and how. *Information Fusion*, 39 :81–95, 2018.
- [FGW10] G. Forestier, P. Gançarski, and C. Wemmert. Collaborative clustering with background knowledge. *Data & Knowledge Engineering*, 69(2) :211–228, 2010.
- [SF12] Robert E Schapire and Yoav Freund. *Boosting : Foundations and algorithms*. MIT press, 2012.