

Apprentissage collaboratif multiparadigme interactif pour l'analyse de séries temporelles

CONTEXTE

La génération en flux quasi continu d'une quantité de plus en plus massive de données interdit dans bien des domaines d'effectuer la phase d'étiquetage de l'apprentissage supervisé. En effet cette tâche ne peut plus être assurée par les experts car trop fastidieuse et chronophage. De plus, elle suppose que les experts disposent déjà d'une définition a priori des classes (nomenclature, ontologie...) pouvant les intéresser. Or, celle-ci n'est pas toujours disponible ou n'est qu'uniquement partielle. Le projet ANR HERELLES⁽⁺⁾ a pour objectif de répondre à ce besoin en proposant une *méthode innovante d'apprentissage collaboratif multiparadigme interactif*, (forme étendue du Federating Learning) permettant d'associer des méthodes supervisées et non-supervisées tout en permettant l'interaction avec l'expert.

Ce projet s'appuie sur deux aspects complémentaires, l'utilisation d'échanges d'information entre les méthodes en s'inspirant de concepts existants, le boosting [3], le co-learning [4] pour le supervisé et le collaboratif [2] pour le non-supervisée, ainsi que l'ajout de contraintes dans ce processus [1].

DESCRIPTION DU SUJET

La personne recrutée devra proposer et définir des mécanismes originaux permettant à des méthodes supervisées et non-supervisées de collaborer de façon efficace pour arriver à un consensus de classification. Les modalités d'échange d'information entre celles-ci devront être précisées. Elle devra aussi définir un protocole d'interaction entre l'utilisateur et les méthodes d'apprentissage à travers l'utilisation de contraintes. Enfin, elle devra implanter concrètement les approches proposées pour permettre le test et la validation de celles-ci. Un ingénieur dédié à la plateforme FoDoMuST l'aidera dans cette tâche.

COLLABORATION ET ENCADREMENT

La personne recrutée sera co-dirigée par Antoine Cornuéjols (AgroParisTech - 50%), spécialiste de l'apprentissage supervisé collaboratif et Pierre Gançarski (ICube - 50%), spécialiste du clustering collaboratif. Elle collaborera activement avec l'équipe SDC d'ICube à Strasbourg et plus particulièrement avec Antoine Saget doctorant (2ième année) travaillant sur ce même sujet, ainsi que Baptiste Lafabregue, co-auteur, avec P. Gançarski, de plusieurs articles sur le clustering sous contraintes de séries temporelles [1]. Elle pourra donc être amenée à s'y déplacer fréquemment (aux frais du laboratoire).

INFORMATIONS GÉNÉRALES

Lieu : Saclay (Campus d'AgroParisTech, 22 place de l'Agronomie, 91120 Palaiseau) ou au choix
Strasbourg (ICube, 300 bld Sébastien Brant 37400 Illkirch)

Durée : Un an

Date : **Au plus tôt**

Salaire : de 2000€/mois à 2300€/mois (net) en fonction de l'expérience.

Contact : Antoine Cornuéjols antoine.cornuejols@agroparistech.fr et Pierre Gançarski, pierre.gancarski@unistra.fr

PROFIL SOUHAITE

- Doctorat en informatique et spécialisé en apprentissage automatique/fouille de données.
- Connaissances solides en Science des Données et plus particulièrement sur les méthodes standards de classification et de clustering. Une première expérience sur l'utilisation de modèles collaboratifs/ensemblistes ou d'intégration de contraintes serait un plus.
- Bonnes compétences en communication verbale (anglais ou français) et écrite (anglais).
- Compétences interpersonnelles et la capacité à travailler individuellement ou en tant que membre d'une équipe de projet.

POUR POSTULER

Les personnes intéressées doivent soumettre (par courrier électronique adressé à antoine.cornuejols@agroparistech.fr) leur curriculum vitae, la liste de leurs publications, une lettre de motivation et les coordonnées de trois références. Les candidatures seront admises jusqu'à ce que le poste soit pourvu. Le poste débutera le plus rapidement possible.

Vu les délais, les personnes étrangères devront être déjà en possession d'un visa et d'un titre de séjour valide.

(+) <https://herelles-anr-project.cnrs.fr/>

- [1] T. Lampert, T-B-H. Dao, B. Lafabregue, N. Serrette, G. Forestier, B. Crémilleux, C. Vrain, P. Gançarski, Constrained distance based clustering for time-series: A comparative and experimental study. *Data Mining Knowledge Discovery*, 32:1663–1707 (2018)
- [2] Antoine Cornuéjols, Cédric Wemmert, Pierre Gançarski, and Younès Bennani. Collaborative clustering: Why, when, what and how. *Information Fusion*, 39 :81–95, 2018.
- [3] R.E. Schapire et al. *Boosting: Foundations and Algorithms*. MIT Press (2012).
- [4] 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.