



CEA LIST

Stage

Co-clustering de séries temporelles

Contexte

Le Commissariat à l'énergie atomique et aux énergies alternatives (CEA) est un organisme public de recherche qui est un acteur majeur de l'espace européen de la recherche et exerce une présence croissante à l'international.

Au sein du CEA Tech, l'institut CEA LIST focalise ses recherches sur les systèmes numériques intelligents. Porteurs d'enjeux économiques et sociétaux majeurs, ses programmes de R&D sont centrés sur le *manufacturing* avancé (robotique, réalité virtuelle & augmentée, contrôle non destructif, vision), les systèmes embarqués (sûreté & sécurité, ingénierie logicielle et systèmes, architectures de calcul), l'intelligence ambiante (capteurs, instrumentation & métrologie, communication & interfaces sensorielles, traitement de données & multimédia). En développant des technologies de pointe dont les applications couvrent les secteurs des transports, de la sécurité/défense, du *manufacturing*, de l'énergie et de la santé, le CEA LIST contribue à la compétitivité industrielle de ses partenaires par l'innovation et le transfert technologique (www-list.cea.fr).

Au sein de l'institut CEA LIST, le stagiaire évoluera dans le Laboratoire de Sciences des Données et de la Décision (LS2D) qui comprend une trentaine de personnes.

Sujet du stage

Le *co-clustering* est une méthode de *machine learning* non supervisée qui a pour objectif d'identifier la structure en blocs homogènes d'un tableau de données à partir d'une classification jointe des lignes et des colonnes. Depuis 1965, ce problème a été développé sous plusieurs variantes mais son intérêt s'est considérablement accru ces dernières années avec l'arrivée de nombreuses applications comme l'analyse de données textuelles, l'analyse marketing, la génomique, les systèmes de recommandations ou bien encore l'étude de données énergétiques. Ce type d'approches organise simultanément les lignes et les colonnes d'un tableau pour découvrir des blocs homogènes alignés pour proposer une lecture simplifiée des données ou/et pour en extraire des caractéristiques (*feature engineering*) utilisées par la suite dans des modèles de *machine learning*. Parmi les méthodes développées, on distingue deux types d'approches : les factorisations matricielles et les modèles probabilistes.

L'objectif du stage est d'étudier le cas particulier du *co-clustering* pour les séries temporelles. Dans ce cadre, de récentes approches proposent de transformer dans un premier temps ces données en fonctions pour prendre en compte la notion d'ordre due à la temporalité. A partir de ces transformées, il est alors possible d'utiliser une méthode de *co-clustering* en l'appliquant soit sur les coefficients linéaires des différentes fonctions soit sur les projetés de ces fonctions. Dans le cadre de ces deux approches, la méthode de *co-clustering* retenue par les auteurs est le modèle probabiliste de blocs latents.

Pendant le stage, l'étudiant sera amené à étudier et tester empiriquement ces deux approches. Il pourra alors compléter son étude par un état de l'art sur le *co-clustering* de données temporelles. Sur le même principe que les deux méthodes déjà développées, il pourra être intéressant de proposer une solution alternative couplant le *signal processing* (transformées de Besoz par exemple) et le *co-clustering*. Le stage pourra être suivi d'une thèse dans laquelle une méthode alternative notamment basée sur la factorisation matricielle pourra être développée.

Mots clés : *co-clustering, machine learning, apprentissage automatique non supervisé, séries temporelles, signal processing*

Environnement et Prérequis

- **Lieu du stage :** Le stage se déroulera au CEA Saclay, dans le bâtiment DIGITEO
- **Durée :** 6 mois. Les formalités nécessaires au recrutement du candidat étant assez longues, il est recommandé de commencer les démarches au moins 3 mois avant le début du stage.
- **Rémunération :** à partir de 600€ bruts selon profil
- **Prérequis :** Le candidat sera en M2 spécialisé en *machine learning* et *signal processing*.
- **Responsables et contact :**
 - Contact : Aurore LOMET, aurore.lomet@cea.fr
 - Antoine Souloumiac, antoine.souloumiac@cea.fr