



**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](http://www-list.cea.fr)).

Au sein de l'institut CEA LIST, le stagiaire évoluera dans le Laboratoire pour l'Analyse des Données et l'Intelligence des Systèmes qui comprend une soixantaine de personnes.

### **Sujet du stage**

La classification croisée (*co-clustering*) 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. Ce procédé identifie alors un couple de partitions en ligne et en colonne. Parmi les méthodes développées, l'approche probabiliste par modèles de blocs latents utilisent des variables latentes pour définir les classes en ligne et en colonne.

Le but du stage est d'étudier l'utilisation de ce modèle de blocs latents pour le cas particulier des données temporelles. Ce type de méthode est particulièrement intéressant dans le cadre de l'étude de consommation énergétique par exemple ou dans l'analyse de signaux pour les comprendre, les résumer mais aussi pour en extraire des caractéristiques pertinentes (*feature engineering*). Une récente approche propose de transformer ces données en fonction puis d'utiliser une FPCA (*Functional principal component analysis*) dans le cadre d'un prétraitement pour appliquer le modèle de blocs latents. Une seconde approche qui considère aussi les signaux comme des fonctions s'appuie sur une transformation du modèle de blocs latent pour données fonctionnelles (FunLBL).

Pendant le stage, l'étudiant devra dans un premier temps faire un état de l'art des méthodes de co-clustering pour les données temporelles. Dans un second temps, il s'agira d'implémenter les méthodes décrites dans le paragraphe précédent pour les tester sur des signaux énergétiques. L'application directe du modèle de blocs latents sera aussi réalisée afin de les comparer et d'identifier les forces et les faiblesses des différentes approches. Le stage pourra être suivi d'une thèse.

**Mots clés** : model-based clustering, FPCA, classification non supervisée, séries temporelles

### **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 fouille de données (*clustering*, EM, SEM) et traitement du signal.
- **Responsables et contact** : Aurore LOMET : [aurore.lomet@cea.fr](mailto:aurore.lomet@cea.fr)