

## Fiche de poste

### 1. Identification du poste

**Intitulé du poste : Post-doc sur le Clustering sous contraintes par ensemble pour l'analyse des séries temporelles**

**Type de contrat : CDD**

**Catégorie hiérarchique FP: A**

**Durée du contrat/projet : 1 an (renouvelable une fois)**

**Date prévisible d'embauche : 01/11/2024**

**Quotité de travail : 100%**

**Composante, Direction, Service : Pôle API, 300 Bd Sébastien Brant, 67400 Illkirch-Graffenstaden**

**Niveau d'étude souhaité : doctorat en informatique**

**Niveau d'expérience souhaité : \_**

**Contact(s) pour renseignements sur le poste (identité, qualité, adresse mail, téléphone) :**

- Baptiste Lafabregue, [lafabregue@unistra.fr](mailto:lafabregue@unistra.fr)
- Nicolas Lachiche, [nicolas.lachiche@unistra.fr](mailto:nicolas.lachiche@unistra.fr)

**Date de publication de l'offre : 01/08/2024**

**Date limite de réception des candidatures : 01/12/2024**

### 2. Projet ou opération de recherche

Les systèmes automatisés d'acquisition de données et l'augmentation des capacités de stockage ont rendu les données de séries temporelles disponibles dans un large éventail de domaines, de l'observation de la terre à l'industrie. Cependant, ces données sont souvent fournies avec des étiquettes insuffisantes ou inexistantes, ce qui empêche l'utilisation de méthodes supervisées. Dans ce contexte, les méthodes non supervisées peuvent s'avérer précieuses pour aider les utilisateurs à extraire des informations, comme l'identification de différents comportements sur une chaîne de production. Néanmoins, lorsqu'il s'agit d'analyser des séries temporelles, ces méthodes présentent plusieurs inconvénients.

Tout d'abord, la diversité des capteurs et des sources utilisés pour générer des données temporelles entraîne une hétérogénéité importante en termes de format, de volume, de qualité et de richesse de l'information. Par exemple, une seule ligne de production peut inclure un grand nombre de capteurs différents, chacun étant limité par l'API de son fabricant. Cette diversité a donné lieu à un large éventail de méthodes de catégorisation pour l'analyse des séries temporelles, par exemple sur la base de métriques élastiques, de la décomposition des fréquences et de l'extraction de modèles, chacune ayant ses propres avantages et limites, qui peuvent également se compléter les uns les autres.

Deuxièmement, les approches de clustering donnent souvent des résultats qui ne correspondent pas aux attentes ou aux intuitions des experts. Cela est particulièrement vrai si l'on tient compte de l'hétérogénéité

susmentionnée des données de séries temporelles. Par conséquent, l'incorporation de certaines connaissances d'experts, même si elles n'englobent pas tout le spectre des classes réelles, peut améliorer de manière significative la qualité des résultats du regroupement. Ces connaissances sont souvent exprimées sous forme de contraintes. Cependant, ces méthodes souffrent souvent de l'impact négatif des contraintes, ce qui entraîne une diminution de la qualité lorsque des contraintes sont ajoutées.

Enfin, il n'est pas raisonnable de demander aux experts de définir toutes les classes dès le début du projet. En effet, il arrive souvent que toutes les classes ne puissent pas être définies sémantiquement avant qu'une analyse de données n'ait été effectuée. Il est plus pratique d'impliquer les experts tout au long du processus en développant progressivement le traitement et l'analyse des données dans un cycle itératif d'interactions entre l'expert et le système d'apprentissage. L'objectif de cette interaction est de combler le fossé entre les résultats générés par les algorithmes et les connaissances thématiques de l'expert. Ce processus est conçu pour rendre les résultats plus compréhensibles pour l'expert.

### 3. Activités

#### ➤ **Description des activités de recherche :**

La tâche principale de ce post-doc est de développer une méthode de regroupement d'ensemble qui s'appuie sur une diversité de points de vue (c'est-à-dire de représentations ou de métriques). Elle utilisera des contraintes données itérativement par l'utilisateur pour sélectionner et combiner les points de vue appropriés. Il devrait en résulter un meilleur regroupement qui est un consensus des points de vue les plus appropriés, en adéquation avec les connaissances de l'expert, afin d'atténuer les effets négatifs potentiels des contraintes. Pour atteindre ce but, nous devons remplir quatre objectifs :

- Sélectionner un sous-ensemble de métriques/représentations existantes suffisamment indépendantes/diverses (nécessaires pour avoir des points de vue complémentaires) pertinentes pour regrouper des séries temporelles ;
- Définir une méthode d'ensemble générique pour obtenir un résultat de regroupement consensuel à partir des points de vue précédemment sélectionnés qui maximisent le respect des connaissances de l'expert ;
- Proposer une méthode générique pour mettre à jour itérativement le regroupement en intégrant les nouvelles connaissances de l'expert en interaction avec ce dernier ;
- Valider l'opérabilité de la méthode en se concentrant sur les données industrielles, en s'appuyant principalement sur une ligne de production de démonstration de l'un de nos partenaires industriels.

#### ➤ **Activités associées :**

Rédiger des articles, présenter le travail lors de séminaires ou de conférences, participer aux activités de l'équipe.

### 4. Compétences

#### ➤ **Qualifications / Connaissances :**

- Doctorat en informatique, avec une spécialisation en apprentissage automatique.
- Solide connaissance des méthodes d'apprentissage automatique. Une expérience dans l'analyse de séries temporelles et/ou la maintenance prédictive serait également appréciée.

➤ **Compétences opérationnelles /savoir-faire :**

Aisance à l'orale (anglais ou français) et à l'écrit (anglais).

➤ **Savoir-être :**

Compétences interpersonnelles et capacité à travailler individuellement ou au sein d'une équipe de projet.

## 5. Environnement et contexte de travail

➤ **Présentation de la composante / unité de recherche :**

L'équipe "Science des données et connaissance" du laboratoire ICube couvre un large spectre de recherche en informatique, plus précisément en intelligence artificielle. Nos activités de recherche se concentrent sur deux thèmes de recherche théoriques : l'apprentissage automatique et les données et la connaissance.

Nous sommes spécialistes de certains types de données et nous avons quelques domaines d'application privilégiés.

Les membres de l'équipe participent à des projets de recherche en collaboration avec d'autres laboratoires de recherche ou des entreprises. Ces activités de recherche s'appuient sur notre plateforme regroupant les logiciels sur lesquels nous travaillons.

➤ **Relation hiérarchique :**

La personne recrutée sera codirigée par Nicolas Lachiche (50%), spécialiste de la fouille de données complexes, et Baptiste Lafabrègue (50%), spécialiste de l'analyse des séries temporelles. Elle collaborera activement avec l'équipe DDC d'ICube à Strasbourg, et plus particulièrement avec Nassime Mountasir, doctorant de 3ème année travaillant sur les problématiques de maintenance prédictive.

➤ **Conditions particulières d'exercice (cf annexe jointe):**

-

**Pour postuler, veuillez adresser CV, lettre de motivation le(s) diplôme(s)  
à l'attention de :**