

Position description

1. Position identification

Title of post : Post-doc on Ensemble constrained clustering for time series analysis

Type of contract : fixed-term contract

Category (A,B or C) : A

Contract/project period : 1 year (renewable once)

Expected date of employment : 01/11/2024

Proportion of work : 100%

Workplace : Pôle API, 300 Bd Sébastien Brant, 67400 Illkirch-Graffenstaden

Desired level of education : PhD in Computer Science

Experience required :

Contact(s) for information on the position (identity, position, e-mail address, telephone) :

- **Baptiste Lafabregue, lafabregue@unistra.fr**
- **Nicolas Lachiche, nicolas.lachiche@unistra.fr**

Date of publication : 01/08/2024

Closing date for the receipt of applications : 01/12/2024

2. Research project or operation

Automated data acquisition systems and increasing storage capacities have made time series data available across a wide range of domains, from earth observation to industry. However, this data is often provided with insufficient or no labels, thus preventing the use of supervised methods. In this context, unsupervised methods can be valuable to help users extract information, such as identifying different behaviors on a production line. Nevertheless, when it comes to time series analysis, these methods face several drawbacks.

First, the diverse nature of sensors and sources used to generate temporal data results in significant heterogeneity in terms of format, volume, quality, and richness of information. For example, a single production line can include a large set of different sensors, each constrained by its manufacturer's API. This diversity has led to a wide range of categorization methods for analyzing time series, e.g., based on elastic metrics, frequency decomposition, and pattern extraction, each with its own advantages and limitations, which can also complement one another.

Secondly, clustering approaches often yield results that do not align with the experts' expectations or intuitions. This is especially true when considering the aforementioned heterogeneity of time series data. Therefore, incorporating some expert knowledge, even if it doesn't encompass the full spectrum of actual classes, can significantly enhance the quality of the clustering results. This knowledge is often expressed in the form of constraints. However, these methods often suffer from the negative impact of constraints, resulting in a decrease in quality when constraints are added.

Finally, asking experts to define all classes at the outset of the project is unreasonable. It is indeed often the case that not all classes can be semantically defined before a data analysis has been carried out. It is more practical to engage experts throughout the entire process as they progressively unfold the data processing and analysis within an iterative cycle of interactions between the expert and the learning system. The goal of this interaction is to bridge the gap between the results generated by the algorithms and the expert's thematic insights. This process is designed to make the results more comprehensible to the expert.

3. Activities

➤ **Description of the research activities :**

The main task of this post-doc is to develop an ensemble clustering method that relies on a diversity of viewpoints (i.e. representations or metrics). It will use constraints given iteratively by the user to select and combine the proper viewpoints. This should result in a better clustering that is a consensus of the most suitable viewpoints, in adequacy with the expert's knowledge, to leverage potential negative effects of constraints. To achieve this goal, we need to fulfill four objectives:

- Select a subset of sufficiently independent/diverse existing metrics/representations (required to have complementary viewpoints) relevant to clusterize time series;
- Define a generic ensemble method to obtain a consensus clustering result from the previously selected viewpoints that maximize the respect of the expert's knowledge;
- Propose a generic method to iteratively update the clustering by integrating new expert's knowledge in interaction with the expert;
- Validate the method operability by focusing on Industry data, mainly relying on a demonstration production line of one of our industry partners.

➤ **Related activities :**

Writing papers, Presenting the work during seminars or conferences, participate to the team activities

4. Skills

➤ **Qualifications/knowledge :**

- PhD in Computer Science, specializing in machine learning/explainability.
- Solid knowledge of Machine Learning methods. Experience in time series analysis and/or predictive maintenance would be also valuable.

➤ **Operational skills/expertise :**

Good verbal (English or French) and written (English) communication skills.

➤ **Personal qualities :**

Interpersonal skills and the ability to work individually or as part of a project team.

5. Environment and context of work

➤ **Presentation of the laboratory/unity :**

The ICube laboratory's "Data Science and Knowledge" team covers a large spectrum of research in computer science, more precisely in artificial intelligence. Our research activities focus on two theoretical research themes Machine learning and Data and knowledge.

We are specialist of some data types and we have a few privileged application domains.

Team members participate to research projects in collaborations with other research laboratories or companies. Theses research activities rely on our platform gathering the softwares we work on.

➤ **Hierarchical relationship :**

The person recruited will be co-directed by Nicolas Lachiche (50%), specialist of complex data mining, and Baptiste Lafabrègue (50%), time series analysis specialist. He or she will actively collaborate with the SDC team at ICube in Strasbourg, and more particularly with Nassime Mountasir, a 3rd-year PhD student working on predictive maintenance issues.

➤ **Special conditions of practice (notice attached):**

To apply, please send your CV, cover letter and diploma to :

Baptiste Lafabregue, lafabregue@unistra.fr and Nicolas Lachiche, nicolas.lachiche@unistra.fr