

# Post-doctoral position in dissimilarity learning for interactive clustering

LITIS lab., Université de Rouen Normandie

from now to 31 dec. 2022

The "Machine Learning" research team at LITIS lab, the computer science lab of the University of Rouen Normandie, is looking for a post-doctoral researcher on a 24-months contract, starting as soon as possible. The position will be financed by the research project SCHISM, jointly held by GREYC (computer science lab at the University of Caen Normandy), CERMN (drug design lab at the University of Caen Normandy) and LITIS.

## 1 Context

The research performed in SCHISM will focus on interactive pattern-mining, and interactive clustering in a chemoinformatics context. Interactive data-mining is a recent research direction that breaks with the older paradigm of specifying parameter settings for algorithms, letting the algorithm run, interpret the results of the operation, and, based on this interpretation, adjust parameter settings to restart the process. Interactive data-mining, on the other hand, proposes partial or preliminary results to the user, collects their feedback, and uses this feedback to bias the mining process going forward.

The overall goal of SCHISM is to develop a robust approach to interactive data-mining that integrates both pattern-mining and clustering, and to deliver a prototype that allows users to launch pattern mining or clustering algorithms, visualize the results, give feedback, and rerun mining operations, which will take the given feedback into account.

## 2 Research description

The general idea of the research work carried out at LITIS lab for the SCHISM project is to rely on dissimilarity representations to both infer relevant clusterings and to propose tools for visualization, analysis and expert feedback under the form of constraints on the clusters. The goal is to define a common representation so that the data/constraints are projected into smaller spaces while preserving the properties of the neighborhood. To do so, a preliminary study

has been carried out the last few months on using Random Forest (RF) models to measure dissimilarities and on using it afterwards for inferring relevant clusterings.

The post-doctoral researcher will be in charge of (i) deepening the use of RF dissimilarities for deriving a relevant clustering and (ii) proposing mechanisms of analysis and interaction with the expert. More specifically, this latter task must meet a two-fold objective, namely allowing the expert to select and analyze (sub-)clusterings in order to propose a feedback, and translating this feedback into the form of dissimilarity constraints in order to adjust the result. This research work will therefore involve becoming familiar with Random Forest methods and interactive clustering.

### 3 Profil and application

The successful applicant will:

1. possess or be on track to complete a PhD in computer science, or applied mathematics with a focus on machine learning or data-mining, Fluency in written and spoken English or French is essential;
2. have strong programming skills (Java, Python, etc.) and in-depth understanding of statistics and machine learning;
3. have a productive publication record;
4. have a strong work ethic and time management skills along with the ability to work independently and within a multidisciplinary team as required. Salary will be in line with European and French guidelines w.r.t. years of research experience. Additional funding is available for travel.

Your application should include:

1. curriculum vitae
2. statement of past research accomplishments, career goal and how this position will help you achieve your goals
3. two representative publications
4. contact information for three references

and should be sent to Laurent Heutte ([laurent.heutte@univ-rouen.fr](mailto:laurent.heutte@univ-rouen.fr)) and Simon Bernard ([simon.bernard@univ-rouen.fr](mailto:simon.bernard@univ-rouen.fr))