Post-doctoral position in interactive non-supervised and semi-supervised data mining

December 5, 2020

The Constraints, Data Mining and Graphs team (CoDaG) is a research group of the GREYC, the computer science laboratory of the University of Caen Normandy. CoDaG is looking for a post-doctoral researcher on a two-year contract, starting on January 1, 2021. The position will be financed by the research project SCHISM, jointly undertaken by the GREYC, the CERMN (Centre d'étude et de recherche sur le médicament de Normandie) and the LITIS (Laboratoire d'informatique, de traitement de l'information et des systèmes) in Rouen.

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

As the name already indicates, CoDaG's research focus is on pattern mining. Current work on interactive pattern-mining uses direct feedback on patterns: the user is presented with a partial selection of patterns then he indicates for individual patterns whether she likes or dislikes them, or he imposes a partial order on the whole partial result set. The post-doctoral researcher will instead exploit indirect feedback to integrate pattern-mining and clustering, and refine the pattern-mining process: users give feedback on a clustering instead of patterns, which will provide information about instances, e.g. instances that must or cannot be grouped together. This piece of information will then be exploited

to mine patterns, e.g. those that discriminate between cannot-be-grouped instances, or those that are common to must-be-grouped ones. Research work will therefore involve becoming familiar and working with concepts from both interactive clustering and interactive pattern-mining.

3 Profil and application

The successful applicant will:

1. possess or be on track to complete a PhD 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. Experience with Linux is a plus.
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

4 Contact

The application should be sent to Albrecht Zimmermann (albrecht.zimmermann@unicaen.fr) and Bertrand Cuissart (bertrand.cuissart@unicaen.fr)