2018-00779 - Post-doctoral Position/Implementing a meta-mining framework for the exploration of complex data [S]

Level of qualifications required: PhD or equivalent
Function: Post-Doctoral Research Visit

Context
Context and positioning:

This project aims to study knowledge discovery systems at a foundational level to understand how the knowledge discovery process should be carried out in view of the data and the mining methods available. There is usually a variety of algorithms to choose from and some criteria to guide our choice. However, there is no clear strategy for combining them that takes into account the relationship between datasets and methods at work. This strategic information should be extracted from data, analysed and evaluated, used for descriptive purposes and reused to guide the strategic combination of mining methods.

Case studies are found in biomedical sciences, in particular, in the “omics” area (genomics, transcriptomics, lipidomics, etc.), where data sets that are complex and highly heterogeneous. In the case of biomedical data a problem could be to identify homogeneous groups of patients with certain diseases and thus contribute to better diagnosis and to efficient treatments by identifying the key drivers of the disease. A challenge is then how to choose the knowledge discovery approaches (symbolic/numerical, unsupervised/supervised) to employ and how to combine them for mining this complex and heterogenous data. There is usually a variety of algorithms to choose from and some criteria to guide our choice. However, there is no clear strategy for combining them that takes into account the relationship between datasets and methods at work. This strategic information should be extracted from data, analysed and evaluated, used for descriptive purposes and to guide the strategic combination of mining methods.

Assignment
Main objectives and Assignments:

Inspired by the frameworks of meta-learning [1,2] through knowledge-based mining and Exploratory Data Analysis (EDA) [3], we aim to defining an operational and reusable framework for hybrid exploratory knowledge discovery. More precisely, the hired post-doctoral candidate is expected to implement an interactive workflow integrating the following modules:

Module I - Cluster analysis. This module aims to identify plausible and meaningful clusters (classes) of individuals from unsupervised heterogeneous data, that may be incomplete and from different sources. We will propose a novel approach based on unsupervised random forests (URF) [4] that was recently improved [5] and that outperforms existing methods in running time while giving similar or better clusterings on numerical data. However, the adaptation to heterogeneous data remains a challenging problem.

Module II - Subgroup discovery. This module implements one of the pattern mining frameworks (namely, Subgroup Discovery (SD) [6,7]) that enables to elicit interesting hypotheses from the data: descriptive patterns and rules. Patterns are thus used as group descriptions and are selected and used according to quality measures and expert preferences. Furthermore, they will be reused in this KD framework in the selection of the cluster and combination of classifiers.

Module III - Strategic combination of mining methods. This module is implemented within a meta-learning framework for a strategic selection [2] and combination of mining methods [8] based on characteristics of the datasets, areas (features) of expertise of the different methods considered as well as corresponding performance measures. We will consider both numerical methods as well as symbolic methods such as pattern mining and formal concept analysis.

Bibliography:


General Information

Theme/Domain: Data and Knowledge Representation and Processing
Information system (BAP E)
Town/city: Villers-lès-Nancy
Inria Center: CRI Nancy - Grand Est
Starting date: 2018-10-01
Duration of contract: 12 months
Deadline to apply: 2018-06-06

Contacts

Inria Team: ORPAILLEUR
Recruiter: Miguel Couceiro / miguel.couceiro@inria.fr

The keys to success

Application deadline

June 6th, 2018 (Midnight Paris time)

How to apply

Upload your CV on jobs.inria.fr; this should be a pdf file of at most 2Mo.
In addition, send the following documents to Miguel Couceiro (miguel.couceiro@inria.fr) and Chedy Raissi (chedy.raissi@inria.fr) in a single pdf or ZIP file:
- CV including a description of your research activities (2 pages max) and a short description of what you consider to be your best contributions and why (1 page max and 3 contributions max); the contributions could be theoretical or practical. Web links to the contributions should be provided. Include also a brief description of your scientific and career projects, and your scientific positioning regarding the proposed subject.
- The report(s) from your PhD external reviewer(s), if applicable.
- If you haven’t defended yet, the list of expected members of your PhD committee (if known) and the expected date of defense (defense, not the manuscript submission).

In addition, at least one recommendation letter from your PhD advisor should be sent directly by their author(s) to Miguel Couceiro (miguel.couceiro@inria.fr) and Chedy Raissi (chedy.raissi@inria.fr)

Applications are to be sent as soon as possible.

About Inria

Inria, the French National Institute for computer science and applied mathematics, promotes "scientific excellence for technology transfer and society". Graduates from the
Main activities

Main activities:
- Propose theoretical solutions for bridging the gap between numerical and symbolical computing
- Develop programs/applications/interfaces
- Write and publish research papers
- Write reports

Skills

The ideal candidate holds a PhD in computer science and/or applied mathematics, is familiar with knowledge discovery techniques (preferably, both symbolic and numerical) and/or decision making tools.

He/she should be acquainted with programming languages, preferably Python, as part of the work entails the implementation of the frameworks developed in course of this post-doctoral work.

He/she should have good English skills.

Benefits package

- Subsidised catering service
- Partially-reimbursed public transport
- Social security
- Paid leave

Remuneration

Salary: 2653€ gross/month