2018-00531 - Deep Learning for Linguistic and Social Network

Contract type: Public service fixed-term contract
Level of qualifications required: PhD or equivalent
Function: Post-Doctoral Research Visit
Level of experience: From 3 to 5 years

About the research centre or Inria department

Grenoble Rhône-Alpes Research Center groups together a few less than 800 people in 35 research teams and 9 research support departments.

Staff is localized on 5 campuses in Grenoble and Lyon, in close collaboration with labs, research and higher education institutions in Grenoble and Lyon, but also with the economic players in these areas.

Present in the fields of software, high-performance computing, Internet of things, image and data, but also simulation in oceanography and biology, it participates at the best level of international scientific achievements and collaborations in both Europe and the rest of the world.

Context

Established in 1967, Inria is the only public research body fully dedicated to computational sciences.

Combining computer sciences with mathematics, Inria’s 3,500 researchers strive to invent the digital technologies of the future. Educated at leading international universities, they creatively integrate basic research with applied research and dedicate themselves to solving real problems, collaborating with the main players in public and private research in France and abroad and transferring the fruits of their work to innovative companies.

The researchers at Inria published over 4,500 articles in 2013. They are behind over 300 active patents and 120 start-ups. The 172 project teams are distributed in eight research centres located throughout France.

The DANTE team is and Inria team located in Lyon hosted by the ENS Lyon and the IXXI Complex System Institute. The goal of DANTE is to study and model the dynamics of large-scale complex networks, e.g. social networks, technological networks, social-communication networks, etc. Building on the advancements of the digital data revolution, our main challenge is to propose generic methodologies and concepts to develop relevant formal tools to model, and analyse the dynamics of dynamical networks and ongoing dynamical phenomena. Our main focus areas are:

- Access and collect data with adapted and efficient tools. This includes a reflexive step on the biases of the data collected and their relations to real activities/application domain.

- Characterise and model the dynamics of complex temporal network structures (a) through the development of novel representation and metrics; (b) through the analysis of sizeable real world datasets; and (c) through large-scale numerical simulation based on novel computational models. We are interested in the structural and temporal properties of dynamical networks and their co-evolution with ongoing dynamical processes.

- Invent original approaches combining graph theory with signal processing. A key point is to capture temporal features in the data, which may reveal meaningful insights on the evolution of the networks.

- Find solutions in the applications domain related to communication networks. In particular, we study the dynamic of the network (available bandwidth, traffic, etc.) and propose solutions to adapt its parameterisation accordingly.

Assignment

Online social platforms provide information about the interaction structure of users forming globally connected networks, where information is exchanged commonly in textual form. While users can be associated to their individual socio-demographic status (SES), linguistic features, or topical interest, they may form well connected communities around commonly interesting subjects. This coupling between text, network, and sociolinguistic communities appears to be a fundamental character of online social networks. However, current approaches usually consider only one side of these correlations, while fail to identify multivariate dependencies between the network, SES, and language. In order to capture these latent multivariate correlations one need to develop novel methodologies which is capable to consider together all the information available on the complex network topology.
The socio-economic, and the textual data. To reach this integration, we aim to take advantage of the recently developed methodological framework of deep neural networks.

The objectives of this postdoc is to explore this context via the development of semi-supervised multi-factorial analysis methods based on deep learning using heterogeneous data. The aim is to infer correlations / patterns that exist between dynamic linguistic variables, mesoscopic structure and social network dynamics and their socio-demographic and socio-economic attributes. We will base our studies on a corpus of 200 Million tweets from 2 Million users that has been collected over a period of 2 years within the DANTE team. In addition to this corpus itself, we combine this textual data with network informations inferred from tweets and user profiles, and socioeconomic informations obtained from census data.

**Main activities**

The successful applicant will work on the following topics:

- Collection, statistical analysis, and mining of large digital datasets recording the blog posting activity and interactions of millions of users on Twitter.
- To development of statistical learning methods to understand the co-evolution of the network structure, language dynamics, and individual variables like SES, location, or demographic variables.
- To carry innovative methods for the representation of results.

**Skills**

Applicants should have a PhD degree in computer science, physics, or related discipline with strong interest in complex networks, social phenomena, and computational linguistic.

Background in complex networks, data analysis, computational modelling is an advantage.

Efficiency in programming, data collection and analysis are required.

Good academic writing and presentation skills in English are required.

There are no teaching obligations but opportunities.

**Benefits package**

- Subsidised catering service
- Partially-reimbursed public transport
- Social security
- Paid leave
- Flexible working hours
- Sports facilities

**Remuneration**

**Gross salary:** 2650 Euros per month