2018-01044 - Postdoctoral Researcher - Personalized machine learning and speech processing on the edge M/F

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

About the research centre or Inria department

The Inria Lille - Nord Europe Research Centre was founded in 2008 and employs a staff of 360, including 300 scientists working in sixteen research teams. Recognised for its outstanding contribution to the socio-economic development of the Nord - Pas-de-Calais Region, the Inria Lille - Nord Europe Research Centre undertakes research in the field of computer science in collaboration with a range of academic, institutional and industrial partners.

The strategy of the Centre is to develop an internationally renowned centre of excellence with a significant impact on the City of Lille and its surrounding area. It works to achieve this by pursuing a range of ambitious research projects in such fields of computer science as the intelligence of data and adaptive software systems. Building on the synergies between research and industry, Inria is a major contributor to skills and technology transfer in the field of computer science.

Context

Inria Lille is seeking a postdoctoral researcher for a new European (H2020 ICT) collaborative project called COMPRISE. The successful candidate will be part of the Magnet team, which gathers 15 researchers (faculty, postdocs, PhD students) in the field of machine learning, with focus on learning from graph-structured data as well as decentralized and privacy-friendly algorithms. The team is very international and English is the working language.

COMPRISE is a 3-year Research and Innovation Action (RIA) aiming at new cost-effective, multilingual, privacy-driven voice interaction technology. This will be achieved through research advances in privacy-driven machine learning, personalized training, automatic data labeling, and tighter integration of speech and dialog processing with machine translation. The technology will be based on existing software toolkits (Kaldi speech-to-text, Platon dialog processing, Tilde text-to-speech), as well as new software resulting from these research efforts.

The consortium includes academic and industrial partners in France (Inria, Netfective Technology), Germany (Ascena, Saarland University), Latvia (Tilde), and Spain (Rooter).

Assignment

The postdoctoral researcher will work on the design and the validation of privacy-friendly speech-to-text based on machine learning techniques. One important aspect of this work will be to able to personalize the learning phase, taking into account both globally and locally trained models, designing so-called semi-decentralized (private) learning algorithms. He/she will address the following research questions:

- How to design semi-decentralized private speech-to-text training algorithm(s);
- How to formalize measure and control the divergence between local and global models;
- How to exploit these algorithms to adapt deep learning-based speech-to-text models to individual speakers in practice.

This research has both a theoretical machine learning component, mainly focusing on the underlying statistical theory, and a practical speech processing component, focusing on the issue of speaker adaptation.

The research and experimentation will be conducted in the Magnet team at Inria Lille, in tight collaboration with the Multispeech team at Inria Nancy and in connection with Saarland University and Rooter. Depending on his/her desires and aspirations, the successful candidate will have the opportunity to join Multispeech for extended periods of time, in order to benefit from the complementary scientific environments offered by the two teams.

This contract is for 2 years and may be renewed up to 2 more years.

Main activities

- Design semi-decentralized privacy-friendly training algorithms
- Exploit them for speaker adaptation of speech-to-text models
- Implement algorithms in a library compliant with the other developments of the consortium
- Assess privacy/confidentiality of information in a semi-decentralized system
- Validate the proposed solutions
- Coordinate with the two other postdoctoral researchers to be recruited by Inria as part of this project, who will be working on formal privacy models and automatic data labeling

Skills

Two alternative profiles are welcome, either:

- Strong background in mathematics, machine learning, statistics and algorithms
- Strong experience with implementation and experimentation, distributed systems, speech processing.

Excellent English writing and speaking skills are required in any case.
Benefits package

- Subsidised catering service
- Partially-reimbursed public transport
- Sports facilities