A propos du centre ou de la direction fonctionnelle

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 Hauts-de-France 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.

Contexte et atouts du poste

This project will be carried out in the MAGNET Team at INRIA Lille. The work will be supervised by Aurélien Bellet (HDR planned in 2020) and Marc Tommasi. Collaborations with teams at the national and international levels are envisioned. At the national level, Joseph Salmon (optimization for machine learning, University of Montpellier), Emmanuel Vincent (privacy-preserving privacy) and Martin Jaggi groups at EPFL, as well as the privacy-preserving data analysis group at Alan Turing Institute London. The candidate will have the opportunity to spend some time abroad in these renowned teams. Privacy is also a key theme in the European H2020 project Comprise (led by E.Vincent) and the French project PRIVIA (led by M. Tommasi) in which several companies are involved and can provide data and specific use-cases.

Mission confiée

This PhD project focuses on the problem of decentralized and privacy-aware machine learning and optimization. We assume that some users (nodes) in the network have some personal dataset, and we would like to learn models through a collaborative protocol based on decentralized optimization of a joint objective function. This joint objective comes in the form of an empirical risk minimization framework where a certain (regularized) loss function has to be minimized and we aim to analyse and propose new decentralized optimization techniques that are communication-efficient and respect privacy constraints. To formally measure privacy, we will rely on differential privacy [7], which is usually enforced by introducing noise and randomization. For each research direction, we aim to prove theoretical guarantees on the trade-off between optimization convergence and privacy. Practical experiments on benchmark datasets will also be conducted during the PhD, for instance on speech and medical data.

Principales activités

The first year will be devoted to the study of the state of the art. Study will cover various convergence proof techniques of (stochastic) gradient descent approaches. The PhD candidate will also make connections with primal-dual and Frank-Wolfe approaches, block coordinate descent, and optimization techniques robust to noise. The first idea will be to exploit the constraint of moderate and sparse communications to leverage standard privacy techniques (such as adding random noise) and assess the benefits in terms of utility-privacy trade-offs. Other data transformations which can contribute to privacy or anonymity, such as quantization, hashing, or averaging, will also be explored. We will consider optimization algorithms beyond classic SGD, such as coordinate descent and screening methods to better scale to high-dimensional problems.

In the second year, the candidate will study another research direction related to locality: decentralized protocols assume communication in a graph where agents only communicate with their direct neighbors. This locality constraint is generally a way to achieve scalable and asynchronous computations. We will study how to optimize this locality in a data-dependent way so as to improve convergence rates in the non-i.i.d. setting. We also aim to analyze its effect from the privacy perspective by characterizing privacy leaks conditioned to the distance from the attacker in the network, building upon recent work. Connections with related questions such as the robustness to malicious peers, on-line learning, and decentralized similarity learning can also be explored. In the third year, we aim at evaluate and improve impact of research results. It is also devoted to the writing of the PhD manuscript. We plan to participate in deployments of privacy-by-design optimization and learning methods to real studies in collaboration with our partners. A privileged direction is to devise algorithms for collaborative analysis in the medical domain, when participants (hospitals, medical structures or even wearable devices) do not directly share their data.

Compétences

A good candidate will have the following skills:

- A good command of English
- A strong background in mathematics
A good knowledge of machine learning, statistics and algorithms
Preferably some knowledge on distributed systems
Some experience with implementation and experimentation

Please follow the instructions given in https://team.inria.fr/magnet/how-to-apply/ to set up your application file.

Avantages
- Subsidized meals
- Partial reimbursement of public transport costs
- Leave: 7 weeks of annual leave + 10 extra days off due to RTT (statutory reduction in working hours) + possibility of exceptional leave (sick children, moving home, etc.)
- Possibility of teleworking (after 6 months of employment) and flexible organization of working hours
- Professional equipment available (videoconferencing, loan of computer equipment, etc.)
- Social, cultural and sports events and activities
- Access to vocational training
- Social security coverage

Rémunération
1st and 2nd year : 1 982€ Gross monthly salary (before taxes)
3rd year : 2085€ gross monthly salary (before taxes)