2021-03391 - PhD Position F/M Learning to Adaptively Attack and Defend Privacy of Machine Learning Systems

Contract type : Fixed-term contract
Level of qualifications required : Graduate degree or equivalent
Fonction : PhD Position
Level of experience : Recently graduated

Context

In his/her journey to the doctoral thesis, the candidate will be supported by AI_PhD@Lille grant, and supervised by Debabrota Basu and Philippe Preux. Debabrota and Philippe are affiliated with the Scool project-team (previously SequeL) of Inria Lille- Nord Europe. As a team, Scool is internationally recognised for developing theories and algorithms for sequential learning and decision making, i.e. in the fields of bandits and reinforcement learning.

The project is expected to simulate the existing and new collaborations with researchers and groups working on data privacy, privacy-preserving machine learning, and reinforcement learning. Existing collaborators in such topics are distributed internationally at National University of Singapore, Grandes Écoles in Paris, and University of Oslo. In future, the candidate will be encouraged to visit the collaborators and work with them. The candidate will also be part of the HumAIn alliance that aims toward studying humane impact of deploying AI.

From the application point of view, Scool is involved in multiple projects that incorporates medical data, agricultural data, and e-commerce. Depending on the future development, we will be interested to deploy such private systems and algorithm for securing such applications involving individual data.

Assignment

The successful candidate will do research at the intersection of data privacy, privacy-preserving machine learning, and reinforcement learning.

The celebrated success of machine learning depends on availability of large amount of data. Use of this enormous amount of data invokes the concern regarding data privacy. This has led to official regulations like GDPR, and scientific developments like differential privacy. Though differential privacy is well-studied in the offline setting, it is still not completely understood for online learning and decision making problems. We aim to explore and understand differential privacy in online learning and decision making, which brings us to reinforcement learning.

The goal of the PhD thesis is to contribute new theory and algorithms that clarifies fundamental understanding of privacy leakage from reinforcement learning algorithms and how to design optimal differentially private reinforcement learning algorithms. The projected research objectives are:

1. Designing attacks using reinforcement learning that can adapt themselves with online algorithms
2. Designing adaptive defense mechanisms that can secure the data used in online machine learning, such as bandits and reinforcement learning.
3. Incorporating both attack and defense in a single framework of reinforcement learning that operates around data privacy
4. Validating effectiveness and efficiency of proposed methods in practical applications

Along with the research papers, we expect that the different phases of the project will lead to open-source software and their deployment for practical applications involving individual data.

Main activities

The successful candidate is expected to:

1. Study the literature of privacy-preserving machine learning and reinforcement learning
2. Design adaptive attacks on online and reinforcement learning systems
3. Using the knowledge of the leaked data and attach to design private reinforcement learning algorithms
4. Developing a unified framework for understanding privacy defenses and attacks of a real-time reinforcement learning system
5. Developing theoretical privacy and utility analysis at each of these phases
6. Deploying algorithms using standard programming languages like Python in order to validate their applicability

The candidate should aim to publish the research results in premier machine learning (AAAI, AISTATS, ICML, IJCAI, NeurIPS) and privacy enhancing technology (PoPETS, IEEE SSP) venues. Also, the candidate is expected to present his/her work, orally in seminars, workshops, conferences, and also beyond academia towards more general audience.

Since the work involves and impacts the digital life of general public, the successful candidate should collaborate in writing scientific articles aiming towards the larger audience.

Skills

The candidate should preferably have the following skills:

- A strong background in mathematics/statistics
- A good knowledge of machine learning, statistics, and algorithms
- Broad interest for differential privacy or data privacy
- Knowledge of programming languages such as Python. C/C++
- Some experience with implementation and experimentation (a plus)
- A good command of English (a plus)

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

In brief, the application of the candidate should include his/her CV, an application letter, (two or more) recommendation letters, and the school transcripts. It is reccommended that the candidate

General Information

- **Theme/Domain :** Optimization, machine learning and statistical methods
  Statistics (Big data) (BAP E)
- **Town/city :** Villeneuve d'Ascq
- **Inria Center :** CRI Lille - Nord Europe
- **Starting date :** 2021-09-01
- **Duration of contract :** 3 years

Contacts

- **Inria Team :** SCOOL
- **PhD Supervisor :**
  Basu Debabrota / debabrota.basu@inria.fr

About Inria

Inria is the French national research institute dedicated to digital science and technology. It employs 2,600 people. Its 200 agile project teams, generally run jointly with academic partners, include more than 3,500 scientists and engineers working to meet the challenges of digital technology, often at the interface with other disciplines. The Institute also employs numerous talents in over forty different professions. 900 research support staff contribute to the preparation and development of scientific and entrepreneurial projects that have a worldwide impact.

The keys to success

A successful candidate should:

- collaborate inside the team, and with the external researchers and engineers if needed
- organise the work systematically
- be keen to learn new theory and algorithms developed in the fast-changing field of ML and privacy
- engage in meetings and discussions regularly

Instruction to apply

Defence Security :
This position is likely to be situated in a restricted area (ZRR), as defined in Decree No. 2011-1425 relating to the protection of national scientific and technical potential (PPST).Authorisation to enter an area is granted by the director of the unit, following a favourable Ministerial decision, as defined in the decree of 3 July 2012 relating to the PPST. An unfavourable Ministerial decision in respect of a position situated in a ZRR would result in the cancellation of the appointment.

Recruitment Policy :
As part of its diversity policy, all Inria positions are accessible to people with disabilities.

Warning : you must enter your e-mail address in order to save your application to Inria. Applications must be submitted online on the Inria website. Processing of applications sent from other channels is not guaranteed.

contacts Debabrota and Philippe while preparing the application.

The deadline for application is 15th April, 2021.

Benefits package

- 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