PhD offer

Distributed Machine Learning for IoT applications

Project Description:

IoT applications will become one of the main sources to train data-greedy machine learning models. Until now, IoT applications were mostly about collecting data from the physical world and sending them to the Cloud. Google's *federated learning* already enables mobile phones, or other devices with limited computing capabilities, to collaboratively learn a machine learning model while keeping all training data locally, decoupling the ability to do machine learning from the need to store the data in the cloud. While Google envisions only users' devices, it is possible that part of the computation is executed at other intermediate elements in the network. This new paradigm is sometimes referred to as *Edge Computing* or *Fog Computing*. Model training as well as serving (provide machine learning predictions) are going to be distributed between IoT devices, cloud services, and other intermediate computing elements like servers close to base stations as envisaged by the *Multi-Access Edge Computing framework*. This approach provides at least three benefits.

1. **Reduce network load**. According to recent estimates, there are 7 billions IoT devices deployed in the world. This number should increase by a factor 3 by 2025. Routing the raw data traffic generated by these devices to a few data-centers will not be feasible. It is required to extract relevant features as close as possible to the locations where data is generated.
2. **Reduce latency**. ML models will be used by IoT devices to take actions in the physical world. Future wireless services for connected and autonomous cars, industrial robotics, mobile gaming, augmented and virtual reality have strict latency requirements, often below 10 ms and below 1ms for what is now called the tactile internet. A key element to satisfy such constraints is to run these services closer to the user, directly on IoT devices. Edge computing also ensures that applications are not disrupted in case of limited or intermittent network connectivity.
3. **Preserve privacy**. Data captured by IoT devices can contain sensitive or private information. Pre-processing at the edge can make sure that sensitive information is removed or aggregated with data from other devices to preserve user's profile.

Research objectives:

The goal of this PhD thesis is to investigate how both learning tasks and prediction services can be effectively distributed across different elements in the network, taking into account computation/communication constraints.

During the project the candidate will:

- Design new distributed learning algorithms with a particular focus on communication constraints;
- Study analytically their performance and guarantees;
- Implement distributed machine learning algorithms using libraries like PyTorch;
- Participate to the activity of Accenture Labs, interact with research and engineering personnel;
- Interact with Inria students and researchers, and participate to teams' scientific.

Hosting groups:

NEO and Epione teams (Inria Sophia Antipolis) - Accenture Labs (Sophia antipolis). The groups are located in the tech park of Sophia Antipolis in the French Riviera.
The PhD will be hired by Accenture Labs with a yearly gross salary between 35 and 40k€.

Required competences:

Competences in probability, statistics, optimization, and mathematical modeling are essential (Master level). Solid programming and IT skills are necessary (Python, bash, version control systems), along with strong communication abilities.

Contacts:

Giovanni Neglia	giovanni.neglia@inria.fr;
Marco Lorenzi	marco.lorenzi@inria.fr;
Laetitia Kameni	laetitia.kameni@accenture.com;
Richard Vidal	richard.vidal@accenture.com

How to apply:

Applications will be evaluated on a rolling basis.
Send to Giovanni Neglia your application containing:

- CV
- Motivation letter
- Degree certificates and transcripts for Bachelor and Master (or the last 5 years if not applicable)
- Master thesis (or equivalent) if it is already completed, or a description of the work in progress, otherwise
- All your publications, if any
- Contact information of potential referrers, possibly the persons who supervised your Master thesis, or internship

References:

- Giovanni Neglia, Gianmarco Calbi, Don Towsley, Gayane Vardoyan, The Role of Network Topology for Distributed Machine Learning, Proc. of the IEEE International Conference on Computer Communications (INFOCOM 2019), Paris, France, April 29 - May 2, 2019
- Industrial Internet Consortium, The Industrial Internet of Things, Volume T2: Analytics Framework
- Jakub Konecny, H. Brendan McMahan, Daniel Ramage, Federated Optimization: Distributed Machine Learning for On-Device Intelligence, arXiv:1610.02527
- Mehdi Mohammadi, Ala Al-Fuqaha, Sameh Sorour, Mohsen Guizani, Deep Learning for IoT Big Data and Streaming Analytics: A Survey, IEEE Communications Surveys & Tutorials, 2018

- Lorenzo Valerio, Andrea Passarella, Marco Conti, A communication efficient distributed learning framework for smart environments, Pervasive and Mobile Computing 41 (2017) 46–68

About Accenture: Accenture is a leading global professional services company, providing a broad range of services and solutions in strategy, consulting, digital, technology and operations. With 459,000 people serving clients in more than 120 countries, Accenture drives innovation to improve the way the world works and lives.

About Accenture Labs: Accenture Labs incubates and prototypes new concepts through applied R&D projects that are expected to have a significant strategic impact on clients' businesses. Our dedicated team of technologists and researchers work with leaders across the company to invest in, incubate and deliver breakthrough ideas and solutions that help our clients create new sources of business advantage. Accenture Labs is located in seven key research hubs around the world: Silicon Valley, CA; Sophia Antipolis, France; Arlington, Virginia; Beijing, China; Bangalore, India; Herzliya, Israel and Dublin, Ireland. The Artificial Intelligence Research and Development group explores new ways of addressing critical business problems by applying leading edge AI techniques, including machine learning, natural language processing, knowledge representation and reasoning.

About Inria: Inria (www.inria.fr), the French national research institute for the digital sciences, promotes scientific excellence and technology transfer to maximize its impact. It employs 2,400 people. Its 200 agile project teams, generally with academic partners, involve more than 3,000 scientists in meeting the challenges of computer science and mathematics, often at the interface of other disciplines. Inria works with many companies and has assisted in the creation of over 160 startups. It strives to meet the challenges of the digital transformation of science, society and the economy.