

## Offre de stage

---

*Sujet : Adaptive models for decentralized consensus algorithms*

*Possibilité de poursuivre sur une thèse*

### **Encadrement**

Stephan Cléménçon (Telecom ParisTech), Igor Colin (Telecom ParisTech)

### **Lieu et dates du stage**

Telecom ParisTech, 46 rue Barrault, 75013 Paris

Date de début du stage : mars/avril 2017

### **Équipe(s) d'accueil de la thèse**

département TSI, équipe Statistiques et Applications (STA)

### **Mots clés**

Decentralized learning, gossip protocols, adaptive algorithms, dynamic networks

### **Sujet détaillé**

The increasing popularity of large-scale and fully decentralized computational architectures, fueled for instance by the advent of the « Internet of Things », motivates the development of efficient optimization algorithms adapted to this setting. An important application is machine learning in wired and wireless networks of agents (sensors, connected objects, mobile phones, etc.), where the agents seek to minimize a global learning objective which depends on the data collected locally by each agent. In such networks, it is typically impossible to efficiently centralize data or to globally aggregate intermediate results : agents can only communicate with their neighbors (e.g., agents within a small distance), often in a completely asynchronous fashion. Standard distributed optimization and machine learning algorithms (implemented for instance using MapReduce/Spark) require a coordinator node and/or to maintain synchrony, and are thus unsuitable for use in decentralized networks.

In contrast, *gossip algorithms* [2, 6] are tailored to this setting because they only rely on simple peer-to-peer communication : each agent only exchanged information with one neighbor at a time. Various gossip algorithms have been proposed to solve the flagship problem of decentralized optimization, namely to find a parameter vector which minimizes an average of convex functions  $(1/n)\sum_{i=1}^n f(\theta; x_i)$  , where the data  $x_i$  is only known to agent  $i$ . The most popular algorithms are based on (sub)gradient descent

[5, 1], ADMM [4] or dual averaging [3, 7]. The main idea underlying these methods is that each agent seeks to minimize its local function by applying local updates (e.g., gradient descent steps) while exchanging information with neighbors to ensure a global convergence to the consensus value.

The goal of the present research internship is to study the setting where a node can choose the neighbors it will exchange information with. Using an adequate model on the communication budget and the control on the network dynamics, the expected result of this internship is an adaptive algorithm providing improved convergence rates along with strong empirical performance.

## ***La Chaire Machine Learning for Big Data***

Le traitement statistique des masses de données convoque à la fois mathématiques appliquées et informatique, à travers une discipline en plein essor : le Machine Learning ou apprentissage statistique.

Créée en septembre 2013 avec le soutien de la Fondation Télécom et financée à hauteur de près de 2 M€ par quatre entreprises partenaires : Criteo, PSA Peugeot Citroën, Safran et BNP Paribas, la Chaire Machine Learning for Big Data est portée par le mathématicien Stéphan Cléménçon, Enseignant-Chercheur, Professeur au sein du Département du Traitement du Signal et des Images à Télécom ParisTech.



Proposant cinq axes de recherche méthodologiques, enrichis par des applications industrielles concrètes, cette Chaire a pour objectif d'animer, en interaction avec ses partenaires, une activité de recherche de pointe en Machine Learning, ainsi que de proposer des programmes de formation.

La variété des données aujourd'hui disponibles (nombres, images, textes, signaux), leur grande dimension et leur volumétrie rendent souvent inopérantes les méthodes statistiques traditionnelles reposant sur le prétraitement humain et un long travail de modélisation. Le Machine Learning vise donc à élaborer et étudier des algorithmes, à vocation prédictive le plus souvent, permettant à des machines d'apprendre automatiquement à partir des données et à effectuer des tâches de façon performante.

Les avancées technologiques, l'omniprésence des capteurs (systèmes embarqués, objets connectés, Internet...) et l'explosion des réseaux sociaux s'accompagnent d'un véritable déluge de données, propulsant les sciences de l'information au centre du processus de valorisation des masses de données. En plus de la collecte et du stockage, l'enjeu est de pouvoir analyser ces données afin d'optimiser les décisions et mettre au point de nouvelles applications.

Au-delà du buzz médiatique dont il fait l'objet, le Big Data est donc un sujet stratégique majeur, au cœur d'enjeux économiques et sociétaux considérables. Son impact est désormais perçu dans presque tous les secteurs de l'activité humaine : de la recherche scientifique à la médecine en passant, entre autres, par la finance, le bâtiment, l'e-commerce, la défense ou les transports.

En savoir plus sur la Chaire, ses axes de recherche, ses activités, ses publications :

<http://machinelearningforbigdata.telecom-paristech.fr>

## ***Profil du candidat***

Etudiant titulaire d'un master 2 recherche

- Apprentissage statistique / reconnaissance des formes
- Optimisation convexe / distribuée
- Bon niveau en programmation (Java, C/C++, Python)
- Bon niveau d'anglais

## ***Candidatures***

Envoyer à [stephan.clemencon@telecom-paristech.fr](mailto:stephan.clemencon@telecom-paristech.fr) et [igor.colin@telecom-paristech.fr](mailto:igor.colin@telecom-paristech.fr) :

- Curriculum Vitae
- Lettre de motivation personnalisée expliquant l'intérêt du candidat sur le sujet (directement dans le corps du mail)
- Relevés de notes des années précédentes
- Contact d'une personne de référence

Les candidatures incomplètes ne seront pas examinées.

## ***Références***

- [1]** P. Bianchi and J. Jakubowicz. Convergence of a Multi-Agent Projected Stochastic Gradient Algorithm for Non-Convex Optimization. *IEEE Trans. Autom. Control*, 58(2):391-405, 2013.
- [2]** S. Boyd, A. Ghosh, B. Prabhakar, and D. Shah. Randomized gossip algorithms. *IEEE Trans. Inf. Theory*, 52(6):2508-2530, 2006.
- [3]** J. Duchi, A. Agarwal, and M. Wainwright. Dual Averaging for Distributed Optimization: Convergence Analysis and Network Scaling. *IEEE Trans. Autom. Control*, 57(3):592-606, 2012.
- [4]** F. lutzeler, P. Bianchi, P. Ciblat, and W. Hachem. Asynchronous Distributed Optimization using a Randomized Alternating Direction Method of Multipliers. In *IEEE CDC*, pages 3671-3676, 2013.
- [5]** B. Johansson, M. Rabi, and M. Johansson. A Randomized Incremental Subgradient Method for Distributed Optimization in Networked Systems. *SIAM J. Optimiz.*, 20(3):1157-1170, 2010.
- [6]** D. Kempe, A. Dobra, and J. Gehrke. Gossip-Based Computation of Aggregate Information. In *FOCS*, pages 482-491, 2003.
- [7]** I. Colin, A. Bellet, J. Salmon, and S. Cléménçon. Gossip dual averaging for decentralized optimization of pairwise functions. *ICML 2016*.