

Offre de stage

Sujet : Limited Memory Stochastic Average Gradient Descent

Possibilité de poursuivre sur une thèse

Encadrement

Robert M. Gower, Pascal Bianchi

Lieu et dates du stage

Telecom ParisTech, 46 rue Barrault, 75013 Paris

Date de début du stage : April 2018

Équipe(s) d'accueil de la thèse

département IDS, équipe Signal, Statistique et Apprentissage (S²A)

Mots clés

Stochastic variance reduced methods, Empirical risk minimization, first order methods

Sujet détaillé

Large scale optimization problems in Machine Learning, most notably the empirical risk minimization problem, has put pressure on the optimization community to design new highly scalable and incremental methods. This pressure first led to the revival of a rather old method from the 1950's, the stochastic gradient descent (SGD) method [1]. SGD and its variants are now widely used in training deep neural, support vector machines and more. Though highly scalable, one needs to tune a sequence of decreasing stepsizes for the method to converge. Furthermore, different problems and application fields require a different sequence of stepsizes to work efficiently. Which in turn means time has to be spent by the user tuning and the resulting sequence is often suboptimal. This issue, among others, has led to the development of stochastic variance reduced gradient methods, that do not require a decreasing sequence of stepsizes, instead, a constant stepsize suffices to guarantee a fast linear convergence. In particular the SAG and SAGA methods [2,3] require no tuning of the stepsize parameter whatsoever, since the convergence theory that supports these methods gives a practical and easily computed stepsize, much to the relief of the user. This advantage, combined with fast linear convergence, has made SAGA/SAGA popular throughout Machine Learning applications

excluding the area of neural nets. This is because SAG/SAGA require storing a large matrix which, in the case of deep neural nets, often exceeds the available cache memory. This internship will directly tackle this issue by developing limited memory variants of the SAG/SAGA inspired on the limited memory BFGS method [4] and on going work by the 1st supervisor [5, 6].

This internship thus has both a practical aim, in developing stochastic variance reduced gradient methods that can better scale and be applied to training deep nets, and will also develop a theoretical underpinning to once again spare the user of hand tuning parameters. As such, the internship is completely aligned with the first Axis of the Chair on big data, which is **Axis 1 - Reinforcement learning and stochastic optimization/simulation**.

Thus the intern will be entrusted with, 1) algorithmic design, 2) developing high level open software package 3) analysing the convergence of a stochastic algorithm.

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.



BNP PARIBAS
La banque d'un monde qui change



SAFRAN
AEROSPACE · DEFENCE · SECURITY



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

- Statistics and Probability
- Optimization course
- A good level in programming (Java, C/C++, Python)
- A high level in English

Candidatures

à envoyer à robert.gower@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.

Référence

[1] [Robbins, H.](#); Monro, S. A Stochastic Approximation Method. *The Annals of Mathematical Statistics*. **22** (3): 400, 1951

[2] M. Schmidt, N. Le Roux, F. Bach. Minimizing Finite Sums with the Stochastic Average Gradient. *Mathematical Programming*, **162**(1):83-112, 2016.

[3] A. Defazio, F. Bach, S. Lacoste-Julien. SAGA: A Fast Incremental Gradient Method With Support for Non-Strongly Convex Composite Objectives. *Advances in Neural Information Processing Systems (NIPS)*, 2014.

[4] Liu, D. C.; Nocedal, J. [On the Limited Memory Method for Large Scale Optimization](#). *Mathematical Programming B*. **45** (3): 503–528, 1989.

[5] Robert M. Gower and Peter Richtarik. [Randomized iterative methods for linear systems](#), *SIAM Journal on Matrix Analysis and Applications* **36**(4), 1660-1690, 2015.

[6] Robert M. Gower, Francis Bach. Sketching and Stochastic Average Gradient methods, (work in progress)