Ecole Doctorale STIC

Sciences & Technologies de l'Information et de la Communication

MINES ParisTech

Université Nice SOPHIA ANTIPOLIS

ED STIC - Proposition de Sujets de Thèse

pour la campagne d'Allocation de thèses 2017

Axe Sophi@Stic :	SystèmesRéseaux|
Titre du sujet :	Statistical Physics Methods for Distributed Machine Learning
Mention de thèse :	Informatique
HDR Directeur de thèse inscrit à l'ED STIC :	Konstantin Avrachenkov

Co-encadrant de thèse éventuel :

Nom :	
Prénom :	
Email :	
Téléphone :	

Email de contact pour ce sujet :	konstantin.avratchenkov@inria.fr
Laboratoire d'accueil :	INRIA

Description du sujet :

Title: Statistical Physics Methods for Distributed Machine Learning

Pre-requisites if any: Good knowledge of Physics and Mathematics

Detailed description: Over the last few years, research in computer science has shifted focus to machine learning methods for the analysis of increasingly large amounts of user data. As the research community has sought to optimize the methods for sparse data and high-dimensional data, more recently new problems have emerged, particularly from a networking perspective that had remained in the periphery. These new directions go beyond sparsity of data and concern the

distributed nature of data sources as well as the computation itself.

We feel that statistical physics methods such as Gibbs sampling [3] and Generalized Potts Model [2,4] are particularly well suited to design light complexity, distributed machine learning methods for the tasks of unsupervised and semi-supervised learning [1].

The student is expected to work on both theoretical and practical aspects of the topic. We intend to employ both mean-field methods [5] and replica method [6] for the analysis of the statistical physics based machine learning algorithms.

References:

[1] Avrachenkov, K., Goncalves, P., Mishenin, A., and Sokol, M.
Generalized optimization framework for graph-based semi-supervised learning.
In Proceedings of SDM 2012.

[2] Blatt, M., Wiseman, S. and Domany, E.
Clustering data through an analogy to the Potts model.
Advances in Neural Information Processing Systems, pp.416-422, 1996.

[3] Bremaud, P. Markov chains: Gibbs fields, Monte Carlo simulation, and queues.
Springer, 2009.

[4] Eaton, E. and Mansbach, R.
A Spin-Glass Model for Semi-Supervised Community Detection.
In Proceedings of AAAI 2012.

[5] Nishimori, H.
Statistical physics of spin glasses and information processing: An introduction.
Clarendon Press, 2001.

[6] Mezard, M. and Montanari, A., 2009. Information, physics, and computation.
Oxford University Press.

English version: