# Discovering Good Peers in Decentralized and Personalized Machine Learning

Aurélien Bellet, Marc Tommasi

October 25, 2016

## Team and contact


- Équipe Magnet, INRIA/CRIStAL: http://team.inria.fr/magnet
- Aurélien Bellet (aurelien.bellet@inria.fr), Marc Tommasi (marc.tommasi@inria.fr)


## Keywords

Machine Learning, Distributed Algorithms, Graph-based Learning, Model Propagation.

## Context

Most machine learning algorithms were initially designed to be executed on a single computer that has access to the entire training data. In order to scale up these methods to large amounts of data, a lot of work has gone into developing distributed versions of these algorithms (see e.g., [1]) and new programming paradigms and architectures (Hadoop, MapReduce, Spark, GraphLab...) to deploy them on clusters of machines. However, these frameworks assume that the distribution of data across the network is fully controlled, which no longer matches the reality of sensors networks, peer-to-peer networks, or more generally any network where data are "naturally" distributed due to data collection mechanisms or privacy concerns — e.g., personal data collected and stored on mobile phones.

In this internship, we are interested in the novel setting of *decentralized learning of personalized models* [4], where a set of learning agents collaborate in a peer-to-peer network. Each agent learns a personalized model according to its own learning objective, based on its own dataset but also through interactions with other agents. The network is represented by an arbitrary connected graph, and nodes can communicate directly only with their neighbors. The personalized models have to be learned collectively while meeting potential constraints (communication, privacy...). Applications include decentralized recommender systems [2]: in order to train a reliable recommender for each user, one should rely on the (limited) ratings made by the user but also on information brought by other peers (e.g., those with similar taste).

## Objectives

This internship will mainly focus on the problem of *discovering good peers*. This important question was not tackled in [4] as the authors simply assume that the network graph reflects the similarity between agents (i.e., neighbors in the network graph have similar objectives). However, in practical scenarios, agents must discover peers that can provide useful information. This requires strategies to reliably estimate the similarity between any two agents' objectives using available information, as well as efficient decentralized protocols to discover and exchange with good peers that may be outside the immediate neighborhood of an agent. For the latter part, protocols to incrementally construct a $k$-nearest neighbor network overlay [3] may be adapted.

The tentative work-plan is as follows:

1. Review the relevant literature on decentralized / personalized learning and decentralized overlay construction protocols.
2. Propose and evaluate some strategies to estimate pairwise similarities between agents as well as protocols for good peer discovery.
3. If time permits, investigate some further questions, such as (i) how to obtain generalization bounds for simple classes of models that depend on the goodness of peers, (ii) what is a good trade-off between "exploration" and "exploitation" in peer discovery, (iii) how to find and exploit clusters of similar users, and (iv) how to deal with privacy constraints.

## Skills

Basics in machine learning, algorithms and complexity, linear algebra and probability.

## References


[1] R. Bekkerman, M. Bilenko, and J. Langford. *Scaling up machine learning: parallel and distributed approaches*. Cambridge University Press, 2011.

[2] A. Boutet, D. Frey, R. Guerraoui, A. Jégou, and A.-M. Kermarrec. WHATSUP: A Decentralized Instant News Recommender. In *IPDPS*, 2013.

[3] M. Jelasity, A. Montresor, and Ö. Babaoglu. T-Man: Gossip-based fast overlay topology construction. *Computer Networks*, 53(13):2321–2339, 2009.

[4] P. Vanhaesebrouck, A. Bellet, and M. Tommasi. Decentralized Collaborative Learning of Personalized Models over Networks. Technical report, arXiv:1610.05202, October 2016.