Efficient Peer Discovery for Decentralized Machine Learning

Aurélien Bellet, Marc Tommasi

September 27, 2017

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 or on clusters of machines and have access to the entire training data. However, these frameworks no longer match 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 [6, 1], 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 for instance decentralized recommender systems [2]: in order to train a reliable recommender for each user, one should rely on the small number ratings made by the user but also on information brought by other peers (e.g., those with similar taste).

Objectives

A key assumption made in [6, 1] is that the the network graph reflects the similarity between agents (i.e., neighbors in the network graph have similar objectives). In many situations, this assumption does not match real-world applications (e.g., when the network is a physical communication network) or only provides a weak approximation (e.g., social networks). This internship will mainly focus on the problem of discovering good peers in a network. This requires strategies to reliably estimate the similarity between any two agents’ objectives using available information, as well as to design efficient decentralized protocols to discover and exchange with good peers that are beyond the immediate neighborhood of an agent.

We mainly focus on the case where agents need to perform a supervised classification or regression task. To estimate the similarity between peers, we will consider several scenarios depending on assumptions on the distribution of the data in each peer and exploit connections to the literature in transfer learning, multi-task learning and domain adaptation. To design efficient protocols to
discover good peers given the similarity, we will adapt existing strategies to incrementally construct network overlays [3, 4] or clusters [5].

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. Investigate some additional questions, such as (i) how to obtain generalization bounds for simple classes of models that depend on the goodness of peers, (ii) how to exploit various assumptions on data distributions, and (iii) how to find and exploit clusters of similar users.

Skills

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

References


