Reading Report

Distributed Coordinate Descent Method for Learning with Big Data

This paper proposes a distributed coordinate method, **Hydra**, which can solve loss minimization problems with big data and utilize the parallel processing power of individual nodes to compute efficiently.

At first, the **Hydra** requires to partition coordinates equally to different nodes or computers and store corresponding coordinates locally. At each iteration, the **Hydra** allows each node choosing a random data set with same cardinality τ from those they own, independently from the others. Then, each node is able to compute locally and apply updates to these coordinates based on a simple closed-form formula.

The **Hydra** is the first distributed coordinate method and different from the before paper which are all parallel but not distributed. Moreover, this paper also gives the exact convergency rate analysis and show how it depends on the data and on the partitioning. However, the computation of the gradient of each node requires a lot communication. What's more, the benefit of using multiple nodes is not obvious, whereas, in big data situations one simply has no other choice but to utilize more nodes to improve the performance.

In the following work, the authors propose a fast version of the **Hydra** called *Hydra* ^{2} for regularized non-strongly convex problem. The main step of *Hydra* ^{2} is a proximal forward-backward operation similar to what is to be done in **FISTA** but only compute the proximal operator for the coordinates in the chosen subset of each node at every iteration. By applying the new choice of stepsize parameters, the convergence rate of *Hydra* ^{2} is improved from $O(1/k)$ to $O(1/k^2)$, compared with the **Hydra**.