Communication-Efficient Distributed Dual Coordinate Ascent

Wang He

Considering the real-world systems, communicating data between machines is vastly more expensive than reading data from main memory. In order to reduce the amount of necessary communication, this paper presents a new communication-efficient framework CoCoA, which uses local computation in a primal-dual setting. However, there always a tradeoff between communication and local computation.

The highlight of CoCoA is that it causes a significant reduction in communication cost comes with only a very moderate increase in the amount of total computation. Besides, it is more general than other coordinate descent algorithm since the inner optimizer of CoCoA can use any dual optimization method. Naturally, the convergence rate is related to the inner optimization. For some desirable algorithm such as SDCA as the local optimizer and assuming smooth losses, the convergence rate is geometric. Moreover, there are not strong assumptions on the data, for example, the data is assumed to be orthogonal between the different workers, which always appears in other research papers.

CoCoA focuses on the linear regularized objective functions. The critical step of this algorithm is how to combine partial results from local computation while avoiding conflict with update simultaneously computed on other machines. It has been proven that the dual form of linear regularized problem can be solved efficiently. In each step, each internal optimizer tries to maximize the dual formulation, only with respect to their own local dual variables, in parallel. The core observation of this paper is that the crucial information each node requires about the other dual variables can be compactly represented by a single primal vector without sending around data or dual variables between the machines. Therefore, according to frame of CoCoA, in each round, only a single update vector needs to be communicated to the master node.

Even though it has dramatically reduced the amount of communication, there are some shortcomings existing in this algorithm. At first, it's not a total distributed algorithm. CoCoA requires a master node to converge all the information together. Moreover, CoCoA is just for convex linear predictors with a convex regularization term and hard to extend its results.