Addressing the straggler problem for parameter server(CSCI 5570 proposal)

authors

November 1, 2018

In our experiments, we have found that there are some straggler problems in the parameter systems. Hence, in this proposal, our group would propose some half-baked ideas. And at the end of this proposal, we will give a schedule about the implementation.

1 Introduction

In the past nine weeks, our group have implemented the baseline of the parameter server system. However, we observe some straggler problems. Under Bulk Synchronous Parallel (BSP) computational model, all workers have to wait until the slowest worker finishes the computation in each iteration. The straggler problem will affect the performance of the overall system, especially when the number of the workers increases. In such situation, the computation resource will be wasted for most of the workers are idle.
There are numerous reasons which lead to the straggler problems, e.g. unbalanced data distribution, hardware heterogeneity, resource contention and so on.

2 Possible solution

The first possible solution is that we try to make the data balance across the workers.
The second solution is that we schedule the number of training data before each iteration. To achieve this goal, we will reduce the number of the training data in the straggler worker.

3 Schedule

1. **Week 1. Literature review.**
2. **Week 2. Implementation.**

3. **Week 3. Testing.**
4. **Week 4. Optimization.**