Scaling Up Asynchronous Stochastic Gradient Descent in Multi-core Machines

ABSTRACT

The abstract goes here.

1. INTRODUCTION

- (Cho) SGD is important for machine learning. Not easy to parallelize SGD because of the nature of "sequential updates".
- (Cho) Hogwild (and other similar algorithms)
- Scalability is an issue when scaling it to many cores (give an example). Explain why.
- Describe our approach.
- Why can we address the scalability issue.
- (Huan) Give some experimental results to show the scalability of our algorithm.
- Paper outline.

2. RELATED WORK (CHO)

Review literature for asynchronous SGD, and other asynchronous updates (coordinate descent). I will write this.

3. BACKGROUND (CHO)

3.1 Memory Models in Multi-core Machines

3.2 Asynchronous Stochastic Gradient Descent

1. SGD updates.
2. Hogwild algorithm.
3. Why it cannot scale up (do we want to explain that Hogwild is better in sparse data but not in dense data)?

4. PROPOSED ALGORITHM

5. IMPLEMENTATION ISSUES

6. EXPERIMENTAL RESULTS

Setup, data, machine, . . .

In this section, we evaluate the empirical performance of our proposed algorithm DASGD with other asynchronous algorithms on multi-core machines. We consider ? real datasets: xxx, xxx, and xxx. Detailed information is shown in Table 1.

We compare include the following implementations into comparison:

1. DASGD: our proposed method.
2. (other variants?)
3. HogWild:
4. PASSCoDe:

We run all the experiments on a machine with 40 cores and. . .

We include the following

6.1 Performance on real datasets

Figure: use all 40 cores; maybe only compare our algorithm with HogWild. For each data, show (1) training loss or objective function (2) prediction accuracy.

6.2 Scaling in Number of Cores

Figure: total number of updates vs time

Figure: Scalability (x-axis: seconds × cores, y-axis: objective function) Use 4, 10, 20, 40 cores. (Maybe for our algorithm and for Hogwild).

7. CONCLUSION AND FUTURE WORK

Table 1: Dataset statistics

Dataset	# training samples	# testing samples	# features	#nonzeros
RCV1	?	?	?	?