Motivation

• Modern machine learning problem has large scale and requires distributed setting for storage and computation.
• Communication between the distributed computation unit becomes the bottleneck of the system’s performance.
• Consider the trade-off between accuracy, computation and communication in distributed learning framework.

Definition

• A family of target function $f \in F$, Loss function $L$ and Data set $D$
• With probability $1 - \beta$ and $\forall f \in F$, if

\[
|E_M[L(f(x))] - E_D[L(f(x))]| \leq \epsilon E_D[L(f(x))]
\]

• Then we call $M$ is the Coreset of $D$

Distributed Coreset Boosting

- Initialize the sampling weight
- Each worker construct a Coreset
- Coresets are sent to master node
- Generate weak learner based on Coreset
- Update sampling weight for coreset
- Send back weak learner
- If the algorithm converge?

- Instead of sending the whole local data set to the master, each worker node selectively sends the Coreset.
- Comparing to random sampling, the master node could learn a better classifier based on coreset.

The sensitivity of the sample

- For each sample $x_i \in D$, we define its sensitivity as

\[
\phi(F, l) = \sup_{f \in F} \frac{l(f(x_i))}{\sum_{j=1}^{|D|} l(f(t_j))}
\]

- The sensitivity is large only if there exists at least one function $f \in F$, such that

\[
l(f(x_i)) > l(f(t_j)), \; \forall j \neq i
\]

- Large sensitivity indicates that for the given function family $F$, the sample $x_i$ has larger loss than any other sample in the data set.

Main Theorem

Suppose the feature $X$ is scaled to $[0, 1]$. Assume $h(x)$ is $\eta$-bounded and the empirical loss for $h^t(x)$ satisfies

\[
L^M_{emp}(h^t) \leq (1 + \beta)(1 - \alpha),
\]

then with probability $1 - \delta$, the output of proposed algorithm could achieve error rate

\[
m_n h \in L^* Err(h) + \epsilon
\]

and converges in $O\left(\frac{1}{\epsilon^2} \frac{1}{\epsilon_0} \right)$ iterations.

Result

- AgnBoost (Chen, Shang-Tse, Maria-Florina Balcan, and Duen Horng Chau. “Communication efficient distributed agnostic boosting.” Artificial Intelligence and Statistics. 2016.)