4 P2P distributed training

Peer-to-peer training with deep neural networks has been under-explored in the literature. First, it is not obvious how to track the progress of training in a peer-to-peer paradigm because there is no global consensus. Second, it is not clear before training how many peers you must interact with to achieve generalization and to benefit from all party’s data. Third, because of the aforementioned issues, it is unclear how to decide the number of epochs of training. In the central parameter server paradigm, the issues above do not exist because it’s assumed that all collaborating parties want to contribute their data to just a single model. However, in the peer-to-peer setting, it is at least clear that all parties would like to contribute their data and training is not bottlenecked by a single server. Peer-to-peer decentralizes training making collaboration more amenable to training over real computer networks where peers may be arbitrarily far away or unavailable.

4.1 Inspiration

A source of inspiration for designing a peer-to-peer distributed training paradigm was inspired by evolutionary optimization algorithms, in particular Neuroevolution of Augmenting Topologies. Evolutionary algorithms are based on the principles from natural evolution where populations collectively search parameter space for approximations to the true gradient and update parameters according to consensus. In NEAT, the population of neural networks splits into niches and speciates so that the networks can select among themselves to search through parameter space. Similarly, in SMPL, parties form cliques to search regions of parameter space collectively.

Another source of inspiration for peer-to-peer over a central server was the performance gain of BitTorrent over centralized file sharing. BitTorrent has shown to scale better than central file sharing approaches as more users join the network. BitTorrent systems scale linearly with the number of users and while central approaches scale quadratically. //Check this

4.2 Decentralized Paradigm

Although the aim of this work is not solely focused on the analysis and comparison of distributed training using a central server versus training peer-to-peer, this section briefly elucidates the pros and cons presented by the literature on this conversation.

Advantages

The main advantage of centralized paradigms is scalability. In the centralized paradigm, the central server can become overwhelmed as more nodes join the computation especially when bandwidth constrained.

\cite{lian2017can}

Disadvantages

4.3 Clique Formation Strategies