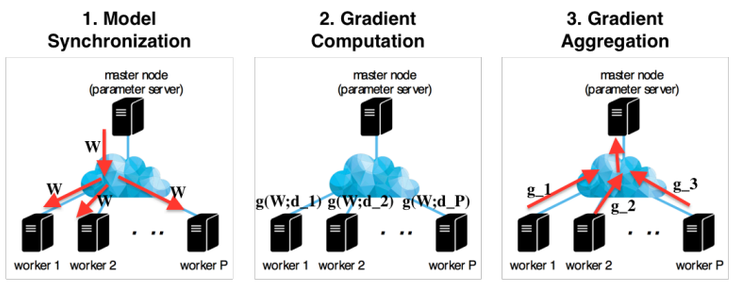
Speeding Up Distributed Machine Learning by Reducing the Network Load

Max Lam, Edward Look, Jesslyn Whittell

How Does Distributed ML Training Work?

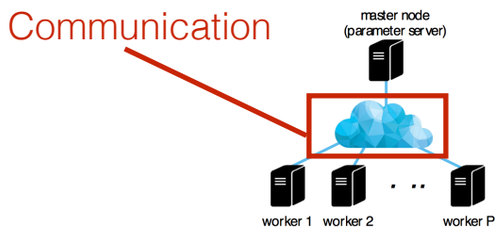


How is Performance?

"Large Scale Distributed Deep Networks" [Dean et al., NIPS 2012]

Not as good as we'd like

What's the Issue?



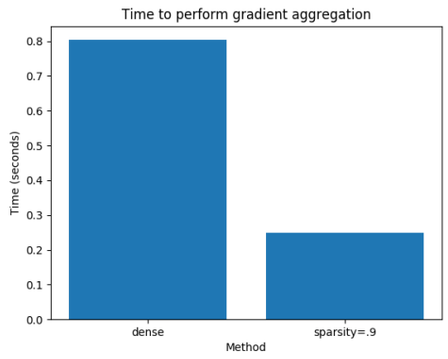
SparsifyGradients Algorithm

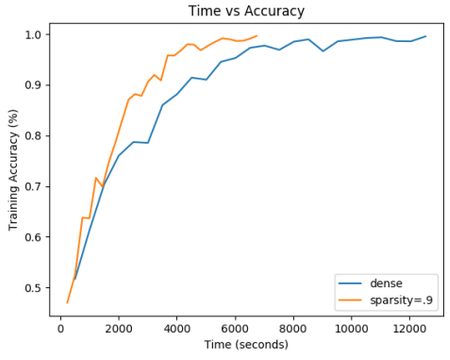
- Drop 90% of small gradient values

Sparsify Gradients Algorithm

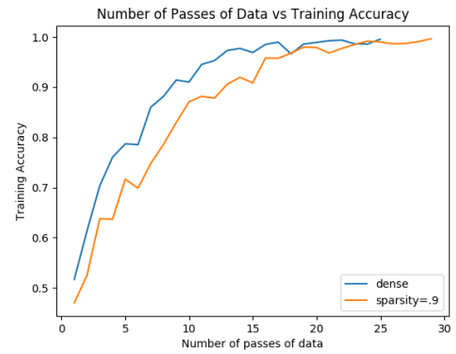
```
def SparsifyGradient(gradient, percentile=90)
    threshold = calculate_percentile_value(abs(gradient), percentile)
    return sparse_format(gradient > threshold)
```

Results

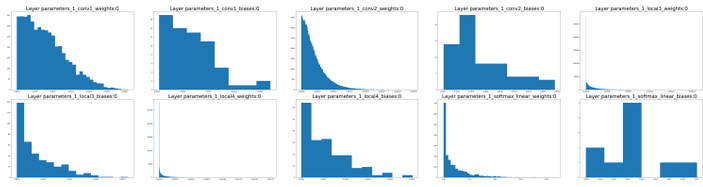




Convergence - Not Too Shabby



Gradient Values Close to 0



Update

As of 4/14/2017, we found out that this idea has been explored by Fikri Aji et al. in https://arxiv.org/abs/1704.05021 with similar results.