PRESENTATION OUTLINE: Determining the susceptibility of MD-GAN networks to nodes containing datasets with different class distributions

Omid Davoudi
School of Computer Science
Carleton University
Ottawa, Canada K1S 5B6
omiddavoudi@cmail.carleton.ca

November 12, 2019

1 Distributed Machine Learning

- Very larger datasets
- More computational power needed
- Globalization of data
- Regulations preventing data movement

2 Distributed Machine Learning - Common Methods

- Parameter Server
- Federated Learning
- Gaia

3 Generative Adversarial Networks (GANs)

- Generative models
- Most commonly used for generating images
- Consist of two neural networks, a Generator and a Discriminator
- Training is tricky, convergence is hard to achieve

4 GAN: Discriminator

- Input shape the same as the data
- Given an input, declares the input as real (from the dataset itself) or fake (artificially generated by the generator network)
- Usually discarded after the training is done

5 GAN: Generator

- Input is a usually a randomly generated latent vector.
- Output is of the same size as the data
- After training is done, it is used to generate new data points.

6 GAN: Training

- A number of fake data points a generated by the generator and the discriminator is trained on a mix of these and some real data from the dataset. It is trained to distinguish between them.
- The weights of the discriminator are frozen and the network is trained end to end in a way that it makes the generator fool the discriminator.
- This process is repeated until (hopefully) convergence

7 MD-GAN

- GAN architecture designed for distributed training
- Main idea is to share parameters for the generator, but have unsynchronized local discriminators
- Assumes each node has a dataset with similar class distributions
- Periodically switches discriminators between nodes to avoid overfitting on local datasets.

8 MD-GAN - Potential vulnerability

- Differences in class distributions between different datasets
- These differences always exist in real life
- Theoretically, they can result in non-convergence
- Switching discriminators between nodes with highly divergent class distributions could result in catastrophic forgetting

9 MD-GAN - Testing the extent of vulnerability

- Proposed test setup: Cluster with at least 4 nodes
- Distribute the data evenly among the nodes
- Train until convergence
- Create skewed datasets each with one class being overrepresented
- Repeat the experiment with increasingly more skewed datasets

10 MD-GAN - Vulnerability results

- Results go here: Convergence time (or lack of convergence) for each dataset

11 MD-GAN - Potential avenue for training speed-up

- Training of the discriminator could happen while the generator is being synchronized
- Training happens using the old generator
- Better utilization of computational power
- Could destabilize the training process or result in non-convergence
- Test done with non-skewed datasets

12 MD-GAN - Speed-up results

- Results go here: Convergence time (or lack of convergence) for using this method compared to not using it

13 Conclusions

- MD-GAN is valid/invalid for real world/skewed datasets
- MD-GAN speed-up did/did not give favourable results

14 References

- References go here...