FID3018 - Opposition Report

Opposition to presentation of Sina Sheikholeslami

Tianze Wang *EECS, SCS*
KTH Royal Institute of Technology
Stockholm, Sweden
tianzew@kth.se

I. Summary

The seminar presents the topic of the effect of dataset configuration in Deep Neural Networks (DNNs) training. The three papers selected to address topics including active learning and parallel training. The seminar is very interesting and engaging with a good discussion on the key contribution of each paper.

II. Choice of Papers

1) Shallue, Christopher J., et al. "Measuring the Effects of Data Parallelism on Neural Network Training." *Journal of Machine Learning Research* 20.112 (2019): 1-49. [1]
2) Chang, Haw-Shiuan, et al. "Active bias: training more accurate neural networks by emphasizing high variance samples." *Proceedings of the 31st International Conference on Neural Information Processing Systems*. 2017. [2]
3) Chitta, Kashyap, et al. "Training Data Distribution Search with Ensemble Active Learning." *arXiv preprint arXiv:1905.12737* (2019). [3]

While the first paper study the effect of batch size on parallel neural network training time, the second and the third paper focus on how training data distribution and data sampling methods will affect the training process in the field of active learning.

Although parallel neural network training and active learning might not seem to be connected at first glance, they are both addressing the topic for training neural networks more efficiently, e.g., train the neural network faster that can also make better predictions. In turn, the topics presented address the challenges of extracting knowledge from massive datasets more efficiently. The combination of parallel training and active learning can be a very interesting research topic that attracts many deep learning researchers.

III. Presentation

The presentation is nicely structured, with proper background information and a clear summary of the motivations and contributions of each work. The brief introduction of active learning gives the audience a proper starting point for understanding the active learning papers. The presentation demonstrates how different strategies of selecting samples for training from a given dataset can affect the speed of training progress (e.g., faster to converge) and the quality of the trained model (e.g., more robust to outliers). The final summary of the presentation also gives a nice aggregated conclusion over the selected papers and what are the interesting aspects to study in the future.

The presentation is easy to follow at a good pace. The bullet points in the slides are very detailed and easy to read through. However, some of the figures presented have small font sizes that might be too condensed to read.

Overall, the speaker offers a nice presentation with an interesting topic that is easy to follow. The presentation also opens up possible ways of combining parallel training and active learning which many researchers will find interesting.

References

[1] C. J. Shallue, J. Lee, J. Antognini, J. Sohl-Dickstein, R. Frostig, and G. E. Dahl, "Measuring the effects of data parallelism on neural network training," *Journal of Machine Learning Research*, vol. 20, pp. 1–49, 2019.
[2] H.-S. Chang, E. Learned-Miller, and A. McCallum, "Active bias: training more accurate neural networks by emphasizing high variance samples," in *Proceedings of the 31st International Conference on Neural Information Processing Systems*, 2017, pp. 1003–1013.
[3] K. Chitta, J. M. Alvarez, E. Haussmann, and C. Farabet, "Training data distribution search with ensemble active learning," *arXiv preprint arXiv:1905.12737*, 2019.

Opposition report for FID3018 (HT20-VT21) : Advanced Course in Data Mining and Analytics.