

FID3018 Opposition Report 2

For Sina Sheikholeslami's session

By Abubakr Sediq Karali

karali@kt.se

The topics of the session covered different problems but somehow it can be grouped under how can we manipulate the dataset configurations for better DNN Training. The topics are very relevant to the course and for me as part of the audience. In the session three main papers were discussed while the last two of them are similar in concept, the first is not:

- Measuring the Effects of Data Parallelism on Neural Network Training (JMLR, 2019, Google)
- Active Bias: Training More Accurate Neural Networks by Emphasizing High Variance Samples (NeurIPS 2017, UMass)
- Training Data Subset Search with Ensemble Active Learning (arXiv, 2020, NVIDIA)

The first paper discusses empirical evaluation of the effect of different batch sizes on the quality of solution and the steps to reach the target accuracy for different models on different datasets. They experimentally characterized the effects of increasing the batch size on training time, as measured by the number of steps necessary to reach a goal out-of-sample error. They studied how this relationship varies with the training algorithm, model, and data set, and find extremely large variation between workloads. They showed that disagreements in the literature on how batch size affects model quality can largely be explained by differences in parameter tuning and compute budgets at different batch sizes. The second paper presented a very interesting idea of a new approach for emphasizing of higher variance that contribute more into training. While the third paper uses active learning also but the goal is not to estimate the importance of each sample during the training it rather searches for the best subset of dataset that is able to represent the whole dataset during the training process to achieve better performance.

The presentation took a little bit more than the required time, the presenter needs to calculate the time required for each slide and plan the presentation accordingly. In general, the talk was organized in an easy to grasp way. However, I would have preferred a common introduction section where he could show a brief introduction to the field, group the papers together, what to expect from the discussed papers and the drawbacks for each.

The presentation highlighted the main contributions in each paper. The presenter gave a very good exposure of the technical details of individual papers and showed solid understanding and could easily go deeper on each. In the QA session, the presenter managed to address the questions put forward by the audience successfully.

Shallue, Christopher J., Jaehoon Lee, Joseph Antognini, Jascha Sohl-Dickstein, Roy Frostig, and George E. Dahl. "Measuring the effects of data parallelism on neural network training." *arXiv preprint arXiv:1811.03600* (2018).

Chang, Haw-Shiuan, Erik Learned-Miller, and Andrew McCallum. "Active bias: Training more accurate neural networks by emphasizing high variance samples." *arXiv preprint arXiv:1704.07433* (2017).

Chitta, Kashyap, Jose M. Alvarez, Elmar Haussmann, and Clement Farabet. "Training Data Distribution Search with Ensemble Active Learning." *arXiv preprint arXiv:1905.12737* (2019).