FID3018 Opposition Report 1

For Tianze Wang session on Device Placement

By Abubakrelsedik Karali
karali@kt.se

The topics of the session had a unified theme of optimizing the device placement for neural networks training. Which is a vital research problem that deals with finding the optimal approach to train neural network on the available hardware resources. The topic is very relevant to the course and for me as part of the audience. In the session three main papers was discussed while the two of them are from the same research group:

- A hierarchical model for device placement
- Spotlight: Optimizing Device Placement for Training Deep Neural Networks
- Post: Device Placement with Cross-Entropy Minimization and Proximal Policy Optimization

The first paper discusses an approach which learns to optimize device placement for training neural networks that have tens of thousands of operations with no need for manual grouping. The method consists of a two-level hierarchical network, in which the first model groups the operations of the graph (the Grouper) and the second model places those groups onto devices (the Placer). The Grouper is a feed forward network which reads in information about each operation and its context within the graph, to predict the group to which that operation should be assigned. The Placer is a sequence-to-sequence model that reads in the embedding of the group and predicts the device placement for that group. In the second paper is based on the claim that using reinforcement learning to learn placement skills by repeatedly performing Monte-Carlo experiments is efficient but due to its equal treatment of placement samples, They proposed new joint learning algorithm, called Post, that integrates cross-entropy minimization and proximal policy optimization to achieve theoretically guaranteed optimal efficiency. So that the training process can complete within the shortest amount of time. The third paper proposed Spotlight, a new reinforcement learning algorithm based on proximal policy optimization, designed for finding an optimal device placement. The design relies upon a new model of the device placement problem: by modeling it as a Markov decision process with multiple stages, we will be able to achieves a theoretical guarantee on performance improvements.

The composition of the presentation in terms of highlighting the interesting aspects of the papers was very good. However, some figures were in low resolution which wasn't very appealing to the eye and not very clear. The presenter gave a very good exposure if the technical details of individual papers and showed solid understanding and could easily go a bit more. In the QA session, the presenter managed to address the questions put forward by the audience successfully. One question I still have for the presenter and in general for this specific field of research is. How are they going to cope with the advances in hardware resources which goes faster and faster and putting the whole research field in danger.

Mirhoseini, Azalia, Anna Goldie, Hieu Pham, Benoit Steiner, Quoc V. Le, and Jeff Dean. "A hierarchical model for device placement." In *International Conference on Learning Representations*. 2018.

Yuanxiang Gao, Li Chen, Baochun Li: Spotlight: Optimizing Device Placement for Training Deep Neural Networks. *ICML 2018: 1662-1670*

Gao, Y., Li Chen and B. Li. "Post: Device Placement with Cross-Entropy Minimization and Proximal Policy Optimization." *NeurIPS* (2018).

FID3018 Opposition Report 2

For Sina Sheikholeslami's session

By Abubakrelsedik 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 is very relevant to the course and for me as part of the audience. In the session three main papers was 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 approaches 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 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 if 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).

FID3018 Opposition Report 3

For Vangjush Komini's session

By Abubakrelsedik Karali
karali@kt.se

The topics of the session covered calibration methods for DNN. While the first two papers deals with post training calibration that requires a validation dataset. The last paper proposed an inference time calibration. The topics is very relevant to the course and for me as part of the audience. In the session three main papers was discussed:

- On calibration of Neural Networks
- Verified Uncertainty Calibration
- Learning for Single-shot confidence calibration in neural network through stochastic inferences

The first paper introduced simple techniques to calibrate a neural network. They observed that depth, width, weight decay, and Batch Normalization are affecting the calibration. They also provide a simple and straightforward recipe for practical settings. The second paper started with the assumption that aperfect calibration does not exist. The observed that that popular recalibration methods are less calibrated than reported, and current techniques cannot estimate how miscalibrated they are. They proposed the scaling-binning calibrator, which first fits a parametric function to reduce variance and then bins the function values to actually ensure calibration. The third paper in contrast learns calibration as a part of inference time through stochastic inference.

The quality of the presentation was excellent. The best in the among the ones we had in the course. The slides are easy to watch, the figures are high in resolution and the equations are clear. However I would have appreciated a lot if the presenter provided a common introduction section where he could show a brief introduction to the field, group the papers together as in slide 31, what to expect from the discussed papers and the drawbacks for each.

The presentation took more than the required time, the presenter needs to calculate the time required for each slide and plan the presentation accordingly. The presentation However was clear. In general, the talk was organized in an easy to grasp way. The presentation highlighted the main contributions in each paper. The presenter gave a very good exposure if the technical details of individual papers and showed solid. I can argue that it was too much technical details and maybe that's why it took so long time. In the QA session, the presenter managed to address the questions put forward by the audience successfully.

Guo, Chuan, et al. "On calibration of modern neural networks." International Conferenceon Machine Learning. PMLR, 2017.

Kumar, Ananya, Percy Liang, and Tengyu Ma. "Verified Uncertainty Calibration."

Seo, Seonguk, Paul Hongsuck Seo, and Bohyung Han. "Learning for single-shotconfidence calibration in deep neural networks through stochastic inferences."Proceedings of the IEEE/CVF Conference on Computer Vision and PatternRecognition.2019.

FID3018 Opposition Report 5

For Vangjush Komini's session

By Abubakrelsedik Karali
karali@kt.se

The topics of the session covered uncertainty in DNN. The topics is very relevant to the course and for me as part of the audience. In the session three main papers was discussed:

- Deep Evidential Regression
- Uncertainty estimation using a single deep deterministic neural network
- Depth Uncertainty in Neural Networks

The first paper deals with uncertainty estimated during the training time. Therefore, sampling in test time is not necessary, also uncertainty is absorbed in a parametric/hierarchical model. They proposed an approach for training non-Bayesian NNs to estimate a continuous target as well as its associated evidence to learn uncertainty. The second paper is on deterministic uncertainty quantification on how to estimate the uncertainty with a single forward path and a single neural network. They used RBF kernel by bringing back centroid update. The third papers proposed a simple to implement and simple to deploy method. they showed that by exploiting the sequential structure of feed-forward networks, we will be able to evaluate our training objective and make predictions with a single forward pass.

The presentation suffered a little from communication bandwidth issues. Figures are not in very good quality slide 5 for example. However, I appreciate a lot that the presenter provided 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 took more than the required time, the presenter needs to calculate the time required for each slide and plan the presentation accordingly. The presentation However was clear. In general, the talk was organized in an easy to grasp way. The presentation highlighted the main contributions in each paper. The presenter gave a very good exposure if the technical details of individual papers and showed solid. I can argue that it was too much technical details which was hard for me to follow as outsider. Also because it there were too many equations maybe that's why it took so long time. In the QA session, the presenter managed to address the questions put forward by the audience successfully.

Amini, Alexander, Wilko Schwarting, Ava Soleimany, and Daniela Rus. "Deep evidential regression." *arXiv preprint arXiv:1910.02600* (2019).

Van Amersfoort, Joost, Lewis Smith, Yee Whye Teh, and Yarin Gal. "Uncertainty estimation using a single deep deterministic neural network." In *International Conference on Machine Learning*, pp. 9690-9700. PMLR, 2020.

Antorán, Javier, James Urquhart Allingham, and José Miguel Hernández-Lobato. "Depth uncertainty in neural networks." *arXiv preprint arXiv:2006.08437* (2020).

FID3018 Opposition Report 5

For Stefanos Antaris's session

By Abubakrelsedik Karali
karali@kt.se

The topics of the session covered different problems temporal GNNs and Reinforcement learning in the session three main papers was discussed while the first two of them are similar in concept, the third is not:

- Streaming Graph Neural Networks
- Learning Temporal Interaction Graph Embedding via Coupled Memory Networks
- End-to-End Deep Reinforcement Learning based Recommendation with Supervised Embedding

In the first paper a new Dynamic Graph Neural Network model was proposed, which can model the dynamic information as the graph evolving. Which overcomes the outcomes of the previous static graphs. The second paper was quite a novel one. They proposed a novel framework named to learn node representations from a sequence of temporal interactions to overcome the drawbacks of learning node embeddings in the context of static, plain or attributed, homogeneous graphs. The third paper is not algorithmic one it mostly towards the architecture of how to apply Reinforcement Learning based Recommendation with Supervised Embedding in production

The presentation took exactly the required time, in an unprecedent achievement in the course i believe. The quality of the presentation was very good. The slides are easy to watch, the figures are high in resolution and the equations are clear in general, the talk was organized in an easy to grasp way. The presenter also provided 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 if the technical details of individual papers and showed solid understanding. In the QA session, the presenter managed to address the questions put forward by the audience successfully.

Ma, Yao, Ziyi Guo, Zhaocun Ren, Jiliang Tang, and Dawei Yin. "Streaming graph neural networks." In *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 719-728. 2020.

Zhang, Zhen, Jiajun Bu, Martin Ester, Jianfeng Zhang, Chengwei Yao, Zhao Li, and Can Wang. "Learning Temporal Interaction Graph Embedding via Coupled Memory Networks." In *Proceedings of The Web Conference 2020*, pp. 3049-3055. 2020.

Liu, Feng, Huifeng Guo, Xutao Li, Ruiming Tang, Yunming Ye, and Xiuqiang He. "End-to-end deep reinforcement learning based recommendation with supervised embedding." In *Proceedings of the 13th International Conference on Web Search and Data Mining*, pp. 384-392. 2020.