



Language
Technologies
Institute

Thesis Defense

Via Zoom | Friday, April 30th | 12:00 pm

**Reducing the Costs to Design, Train, and Collect Data for
Neural Networks with Combinatorial Optimization.**

Hieu Pham

Abstract

The success of modern deep learning algorithms owes itself to the steady effort in scaling up neural models and their training datasets. This combined amazing effort from many research groups have enabled neural network practitioners to train increasingly larger models on increasingly larger datasets, and obtain increasingly better results. Despite the success of large models learning on large datasets, their universal adoption is hindered by their immense expenses. For instance, in 2020, training GPT-3, which is not the largest model for natural language understanding, costs a staggering amount of 4.6 million dollars. While this expense comes mostly from the computations required to train the model, and this cost will eventually go down as better technology becomes available, the same statement does not hold for collecting large training datasets. Due to their expenses, large models and large datasets gradually become a privilege of only corporations with affluent resources.

In this thesis, I present a family of methods to reduce the expense of deep learning models in three facets. This thesis has two key contributions. The first contribution is the Neural Combinatorial Optimization algorithm (NCO), which is the first algorithm that requires no annotated training data yet can still train a recurrent neural network to obtain nearly optimal solutions for certain combinatorial optimization problems, such as the Traveling Salesman Problem. The second contribution is the novel insight that designing, executing, and obtaining training data for neural networks can be formulated as combinatorial optimization problems. Thanks to this insight, I show that by formulating a task of concern as the right combinatorial optimization problem and applying NCO or its variants, I can significantly reduce the expenses of neural networks.

https://drive.google.com/file/d/1XRCA3C9n4D2FyBSjrUp2bqi4j_z1TByi/view

COMMITTEE:

Yiming Yang, (co-chair)

Quoc V. Le, (co-chair)

Samy Bengio,
(Google)

Chris Dyer
(CMU & DeepMind)

Barnabas Poczos,
(CMU & DE Shaw)