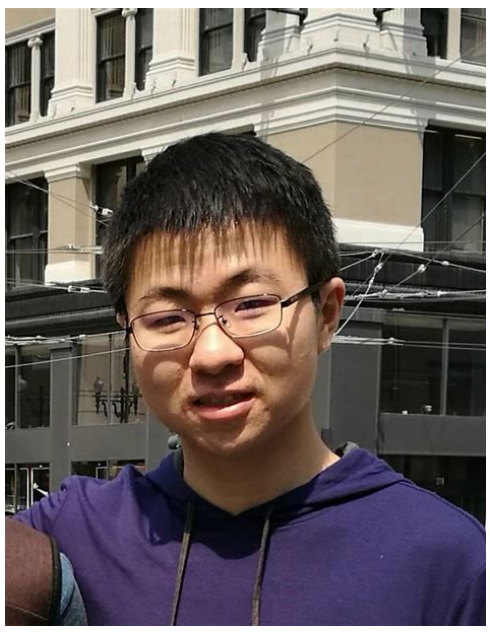
Language Technologies Institute

Thesis Proposal

Via Zoom| April 8, 2021| 9:00 am



COMMITTEE:

Yulia Tsvetkov,
(Co-chair)

Emma Strubell,
(Co-chair)

Graham Neubig

Orhan Firat,
(Google AI)

Improving Robustness and Efficiency of Transferable Machine Learning

Zirui Wang

Abstract

Traditional machine learning paradigm of training a task-specific model on one single task has led to state-of-the-art performance in many tasks (e.g. computer vision and natural language processing). To enable wider applicabilities of machine learning models, transfer learning aims to adapt knowledge learned from source task(s) to improve performance in other target task(s).

However, existing transfer learning paradigm remains a black box, such that we have limited knowledge of its potential limitations, underlying mechanism and solutions for more intelligent transfer. In particular, when transferring knowledge from a less related source, it may inversely hurt the target performance, a phenomenon known as negative transfer. Nonetheless, the root of negative transfer is ill-defined, and it is not clear how negative transfer affect models' robustness and sample-efficiency. In this thesis, with the goal of thoroughly characterizing and addressing negative transfer in machine learning models, we carefully study negative transfer in popular vision and NLP setups, glean insights on its causes, and propose solutions that lead to improved robustness and sample-efficiency. We first conduct systematic analysis of negative transfer in state-of-the-art transfer learning models. We formally characterize its conditions in both domain adaptation and multilingual NLP models, and demonstrate the task conflict as a key factor of negative transfer. Then, we propose two methods to enhance the robustness of transferable models by resolving the aforementioned task conflicts with better-aligned representations and gradients. Finally, we further show that addressing negative transfer can also largely improve the sample-efficiency of transfer learning. We propose methods to train transfer models using explicit meta objectives, such that they can achieve strong transfer performance with less training samples and/or training time.

https://www.dropbox.com/s/f3xyl7w60nbse8e/Thesis_Proposal.pdf?dl=0