Research Statement

Brief Introduction

I have had a sustained interest in trustworthy AI and out-of-distribution generalization. Specifically, I have spent nearly a year as a research intern in Professor Peng Cui's laboratory at Tsinghua University, actively engaging as a key participant in two projects: one focused on stable learning and the other on domain generalization.

Detailed Works

Stable Learning

Methods in stable learning aid in better model generalization to unseen distributions, among which a pivotal approach is sample-reweighting. This method decouples features to eliminate interference from spurious features. However, the expansive solution space often exhibits instability, particularly with a limited sample size. We introduced an additional loss term during the process of learning sample weights, emphasizing weights for poorly performing samples in validation. This addition notably stabilized and enhanced the method.

In this work, I was responsible for translating the senior researchers' conceptual ideas into executable code for experimentation, organizing data, and creating visual representations. This experience exposed me to the fundamental workflow of research and familiarized me with the utilization of Python and PyTorch.

Domain Generalization

Diverging from the conventional model optimization perspective, our focus centered on data acquisition process, especially in scenarios with high data collection costs. Given limited budget and a small number of unlabeled samples from target domain, which domain should be additionally sampled to improve the transferability of backend models? Across multiple datasets, we validated the presence of combinatorial effects and proposed an framework called domain-wise active learning, guiding the data collection process in scenarios involving multiple source domain adaptation.

In this work, I not only independently conducted the experiments but also contributed some original ideas. This experience significantly enhanced my proficiency in scientific research methodologies.

The common thread between these two projects lies in their shared focus on the model's generalization capabilities under distribution shift. They collectively underscore my broad interest in machine learning, showcasing a spectrum of skills—from optimizing models to refining data collection methodologies and a strong will to build trustworthy and robust machine learning models.

Future Expectations

One area that has consistently captivated my attention is the application of AI in high-risk scenarios, such as biomedicine, autonomous driving, and the financial industry. In these domains, the integration of machine learning promises immense assistance. However, model failure on unseen distributions can result in substantial losses. That's why I am particularly concerned with the robustness and transferability of machine learning, seeing this as a prerequisite for AI to evolve into AGI. I believe achieving this is a formidable challenge and a crucial avenue for future research in AI.