Chen Institute Retreat 2023

Presenter: Rafal Kocielnik

Poster title: Can You Label Less by Using Out-of-Domain Data? Active & Transfer Learning with Few-shot Instructions

Abstract: Large Language Models (LLMs) have been adopted for many problems involving sequences (e.g., language modeling, decision making). In domains where unlabeled data is rich, but labeling is scarce or expensive, LLMs can learn the general properties of the data and later be adapted for the task using limited annotation effort. Unfortunately, existing transfer and active learning approaches meant to reduce annotation effort require fine-tuning, which suffers from overfitting to noise and can cause domain shifts with small sample sizes. In this work, we explore the application of LLMs in the domain of social media, where obtaining high-quality labels for custom dimensions of toxicity and social bias is challenging and labor-intensive. We propose a novel Active Transfer Few-shot Instructions (ATF) approach which requires no fine-tuning. ATF leverages the internal linguistic knowledge of pretrained language models (PLMs) to facilitate the transfer of information from existing pre-labeled datasets (source-domain task) with minimum labeling effort on unlabeled target data (target-domain task). Our strategy can yield positive transfer achieving a mean AUC gain of 10.5% compared to no transfer with a large 22b parameter PLM. We further show that annotation of just a few target-domain samples via active learning can be beneficial for transfer. Finally, we find that not all transfer scenarios yield a positive gain, which seems related to the PLMs "familiarity" with the target-domain task.