Robust Machine Learning via Sufficient Invariant and Causal-Aware In-Context Learning

Taero Kim, Hoyoon Byun, Subeen Park, SungJun Lim, Gyeongdeok Seo, Taero Kim, Jihee Kim, **Kyungwoo Song**

Yonsei University

Machine learning models often underperform when faced with distribution shifts between training and testing data. To enhance robustness, we introduce two approaches. First, we propose the Sufficient Invariant Learning (SIL) framework, a novel method designed to capture a comprehensive set of invariant features across diverse environments. To support SIL, we develop Adaptive Sharpness-aware Group DRO (ASGDRO), an algorithm that identifies flat minima to ensure effective generalization across domains. Additionally, we construct a benchmark dataset specifically tailored to evaluate the sufficiency and diversity of invariant feature learning under various distribution shifts. Second, we present Causal-aware In-Context Learning (CCL), which utilizes VAE-based causal representation learning to select task-relevant examples for in-context learning. Unlike traditional methods that rely on superficial linguistic similarity, CCL focuses on causally relevant features, improving performance in out-of-distribution scenarios. Our theoretical and empirical results confirm that SIL and CCL significantly enhance model resilience to distributional changes, ensuring reliable performance across a wide range of challenging environments.