MolKD: Distilling Cross-Modal Knowledge in Chemical Reactions for Molecular Property Prediction

Liang Zeng
Institute for Interdisciplinary Information Sciences (IIIS)
Tsinghua University
zengliangcs@gmail.com

Abstract

How to effectively represent molecules is a long-standing challenge for molecular property prediction and drug discovery. This paper studies this problem and proposes to incorporate chemical domain knowledge, specifically related to chemical reactions, for learning effective molecular representations. However, the inherent cross-modality property between chemical reactions and molecules presents a significant challenge to address. To this end, we introduce a novel method, namely MolKD, which Distills cross-modal Knowledge in chemical reactions to assist Molecular property prediction. Specifically, the reaction-to-molecule distillation model within MolKD transfers cross-modal knowledge from a pre-trained teacher network learning with one modality (*i.e.*, reactions) into a student network learning with another modality (*i.e.*, molecules). Moreover, MolKD learns effective molecular representations by incorporating *reaction yields* to measure transformation efficiency of the reactant-product pair when pre-training on reactions. Extensive experiments demonstrate that MolKD significantly outperforms various competitive baseline models, *e.g.*, 2.1% absolute AUC-ROC gain on Tox21. Further investigations demonstrate that pre-trained molecular representations in MolKD can distinguish chemically reasonable molecular similarities, which enables molecular property prediction with high robustness and interpretability.

AI for New Drug Modalities at NeurIPS 2024.