



Shift-Robust Molecular Relational Learning with Casual Substructure

⌵ Domain	Molecular
⋮ tag	
⌵ Conference / Journal	KDD
≡ Publish year	2023
📅 정리 날짜	@2024년 2월 10일

Summary

Background & Motivation

Molecular relational learning

- Predict interaction behavior between molecular pairs
 - e.g. 약이 어떤 하위 물질로 분해될지
 - drug combination이 어떤 부작용을 일으킬지
 - 화학물질이 어떤 색을 띠 지
- 기존의 trial-error 기반은 한계 있음
 - 모든 조합 실험하기 뻥셈
 - Lab 환경은 error에 취약
 - 인간에 실험필요한 경우 위험

- ⇒ GNN이 molecular property prediction에서 대성공함
- 하지만, GNN의 real-world 적용엔 redundant/noise 등으로 useful info capture 어려움
 - 최근 연구는 그래프의 기능과 상당한 관련성 가진 core substructure 발견에 포커스
 - 다른 원자에 관계없이 어떤 core substructure 가지는 경우 chemical property 결정하기도
 - 화합물의 성질이나 반응 예측을 용이하게 함
 - 이 core substructure은 relational learning과도 연관이 있음
 - Core substructure learning의 challenge는 data 수집과정의 불예측성으로부터 기인한 bias
 - meaningless substructure가 label과 우연한 연관성 있을 수 있음
 - 관련 없는데 shortcut 제공하여 core substructure 확인 없이 학습하는 경우
 - Shortcut에 의지하는 경우 Generalization 어렵게 함
 - Molecular interaction prediction
 - 화학 반응으로 생성된 분자들의 성질을 학습
 - 새 분자 디자인/발견할 때 필수적
 - e.g. Delfos: recurrent NN과 attention mechanism으로 solvation free energy 예측
 - CIGIN: 분자를 그래프 구조로 모델링하여 solvation free energy 예측
 - co-attention map 활용: pairwise importance
 - DDI
 - combination of drugs의 부작용 예측
 - 기존 ML기반: drug fingerprint의 similarity 비교 등
 - MHCADDI: 그래프 co-attention mechanism
 - SSI-DDI: 각 drug의 substructure간의 co-attention
 - MIRACLE: drug간의 link prediction: multi-view graph
 - 기존 방법들의 한계

- casual substructure가느이 causal relationship 고려 X
- specific task에만 타겟해 일반화 어려울 것

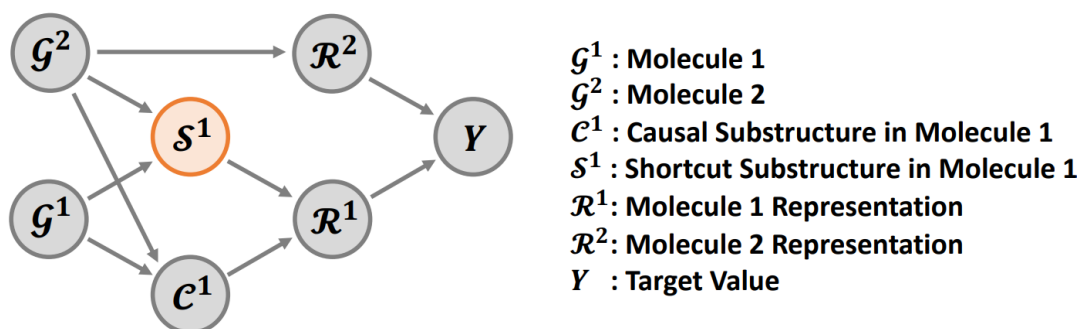
SCM 기반

- Biased graph에서 casual substructrue을 SCM기반으로 찾아 학습하는 방법 많이 제시됨
 - e.g. rationale extraction, graph classification, DIR, CAL
 - DIR: interventional rist간의 variance 줄여 causal substructure 찾아내 GNN prediction에 invariant explantion 제시
 - CAL: 다양한 bias level에 따라 coree substructure 학습해 synthetic graph classification
 - DisC: biased color background에서 causal substructure 찾아냄

Methodology

CMRL

- Robust to distributional shift in molecular relational learning by detecting the core substructure that is causally related to chemical reactions
 - 분자과학의 domain knowledge 기반 인과 relationship 가정하여 SCM(Structural Causal Model) 구축해 variable간 relationship 표현
- causal substructure를 분자 자체 외에도 pair molecule 고려해 결정
 - reveals the causal relationship between a pair of molecules \mathcal{G}^1 & \mathcal{G}^2
 - SCM기반으로 molecular relational learning을 위한 conditional intervention framework 제시



- \mathcal{C}^1 의 intervention space가 pair molecule \mathcal{G}^2 에 conditioned됨
 - 이 confounding effect를 conditional intervention framework로 제거
- target variable Y 에 대해 발견한 causal feature \mathcal{C}^1 의 causal effect를 최대화하여 두 분자간 상호 작용을 예측
 - shortcut feature를 masking하여 \mathcal{G}^2 에 영향받는 \mathcal{S}^1 으로부터 casual substructure \mathcal{C}^1 을 찾음

Questions