Graph Neural Networks in Catalyst Design:
Predicting Activity and Selectivity Under Varied Environmental Conditions

Internship Idea Proposal

Presented by Haejune Lee

Catalyst research

Understand how catalysts work at a molecular level and design new catalysts

- Identifying the **active sites** where reactions occur.
- **Understanding the mechanism** by which catalysts lower the activation energy of reactions.
- **Developing materials** that can perform under desired conditions.

Catalyst research on various fields

- Drug synthesis
- Biocatalysis
- Gas sensors
- Polymer Production
- Chromophores

Traditional catalyst research

The objective of catalyst research

To develop materials that are not only active, meaning they can catalyze a reaction efficiently,

but also selective, meaning they preferentially catalyze the formation of a desired product under a certain condition

The process of catalyst research

Design → Synthesis → Charaterization → Testing → Optimization

Challenges

- **Trial and error: High cost and time for development**
- **Complexity of catalytic processes**
- Scalability and reproducibility
- Durability and deactivation
- Safety

ML application on catalyst research

Black box methods

Increased prediction accuracy, but **hard to interpret how decisions are made**

Molecular Dynamics and Machine Learning in Catalysts (Liu et al.) : AB initio molecualr dynamics and reactive force field molecular dynamics를 이용해 학습

GNN applications

Graph neural networks for materials science and chemistry (Reiser at al.) : Leveraging GNN for handling graph structured data like molecules

Conditional Graph Information Bottleneck for Molecular Relational Learning (DSAIL) : Identify and utilize critical substructures within molecular graphs rather than relying on entire molecular graphs

Shift-Robust Molecular Relational Learning with Causal Substructure (DSAIL) : Adjust for distributional shifts by focusing on causal features

Application

Consideration on the characteristics of catalyst

- **Environment condition**: Catalysts can exhibit significantly different behaviors under varying environmental conditions

- **Multiple active sites**: Capture influence of multiple surface structures

- **Integrate with quantum mechanical calculation**: Many catalyst properties are determinded by QM effect in atomic level

Challenges of this research proposal

Data availability : Obtaining comprehensive datasets that cover a wide range of catalysts, reaction conditions, and performance metrics can be challenging

Model construction : Model should catpture the complex nature including the effects dynamic behavior under conditions

Domain knowledge : Need to deepen understanding in material sciences as well as DNN

Open Catalyst datasets

- Open Catalyst 2020 (OC20)
  - Scripts to download and preprocess the data
  - Structure to Energy and Forces (S2EF)
  - Initial Structure to Relaxed Structure (IS2RS) / Relaxed Energy (IS2RE)
  - Relaxation Trajectories
  - Bader charge data
  - OC20 metadata
  - Changelog
  - License and bibtex
- Open Catalyst 2022 (OC22)
  - Structure to Total Energy and Forces (S2EF-total)
  - Initial Structure to Relaxed Structure (IS2RS) / Relaxed Total Energy (IS2RE-total)
  - Relaxation Trajectories
  - OC22 metadata
  - License and bibtex

Internship recall

Thank you

References

Molecular Dynamics and Machine Learning in Catalysts (Liu et al.)

Graph neural networks for materials science and chemistry (Reiser et al.)

Conditional Graph Information Bottleneck for Molecular Relational Learning (DSAIL)

Shift-Robust Molecular Relational Learning with Causal Substructure (DSAIL)

Interpretable machine learning for knowledge generation in heterogeneous catalysis (Esterhuizen et al.)

The Open Catalyst 2022 (OC22) Dataset and Challenges for Oxide Electrocatalysts (Tran et al.)

Bridging the complexity gap in computational heterogeneous catalysis with machine learning (Mou et al.)

Graph Neural Networks for Learning Molecular Excitation Spectra (Singh et al.)