Graph Neural Networks in Catalyst Design:
Predicting Adsorbtion Energy Under Varied Environmental Conditions

Research Idea Proposal

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 on molecular science

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 setting performance metrics can be challenging

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

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

Building a model to screen candidates

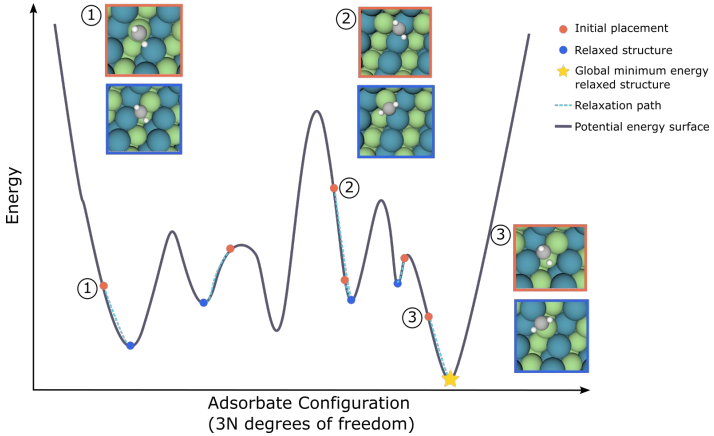
Predicting adsorption energy : How strongly a molecule referred to as the adsorbate binds to the catalyst's surface. The rate of the chemical reaction is correlated with adsorption energies.

Glass-box model : Explanation of why such suggested candidates are chosen should be given.

Reaction condition consideration : Predict how the adsorption energy is changes throughout the conditions change.

Open Catalyst Challenge

Binding energy prediction: Predicting lowest local/global relaxed energy of adsorbate-surface combinations using open catalyst dataset



Global minimum binding energy

Jun 26, 2023: Competition announced

Sep 20, 2023: Challenge test dataset release

~~Oct 13~~ Oct 20, 2023: Submission deadline

Oct 30, 2023: Winners notified

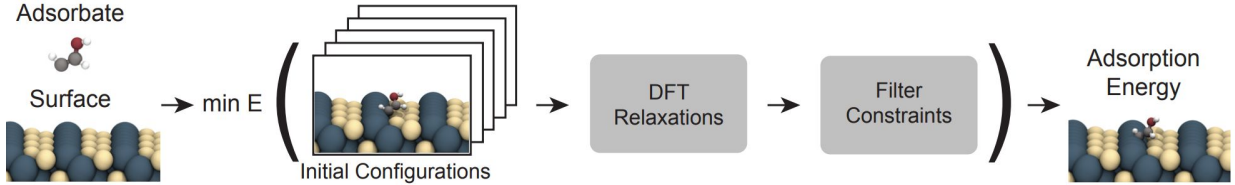
Nov 17, 2023: Presenters' slides and videos due

Dec 16, 2023: Results Presentation @ NeurIPS AI for Science Workshop 2023

AdsorbML: A Leap in Efficiency for Adsorption Energy Calculations using Generalizable Machine Learning Potentials

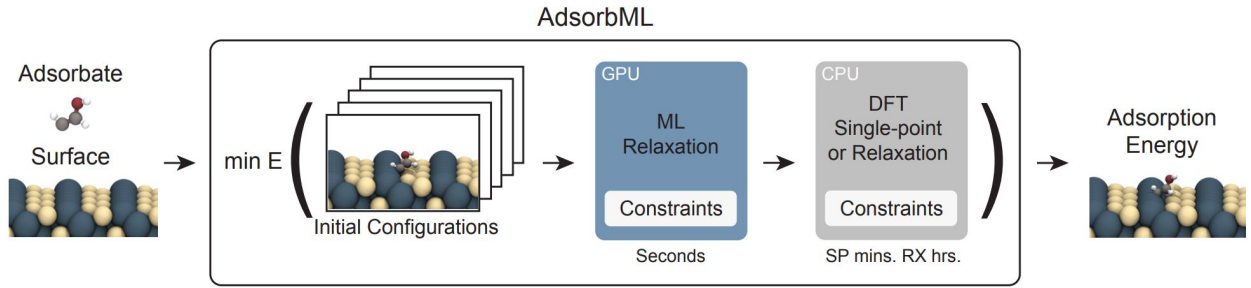
Previous energy calculation : Density Functional Theory(DFT)기반 energy calculation

1. 각 configuration마다 per-atom forces를 iteratively calculate하여 local optimization → local energy minimum : O(N^3)
2. Expert intuition, heuristic approaches



Hybrid approach : DFT와 ML을 같이 사용해 accuracy와 efficiency를 어느정도 충족

AbsorbML : ML 모델이 relaxation 파트를 수행해서 best k 에 대해 DFT 수행



AdsorbML: A Leap in Efficiency for Adsorption Energy Calculations using Generalizable Machine Learning Potentials

Limitation

1. Noise : Implicit or explicit가 ML energy prediction 정확도 낮춤
2. Local optimization만 수행: 전체 반응성 예측 위해서 Global optimization 필요함

Brute force로 local optimization 반복 수행해 global 찾음

이거보다 더 나은 approach 있을 것

Spherical Channels for Modeling Atomic Interactions

원자를 node, 결합을 edge로 한 GNN

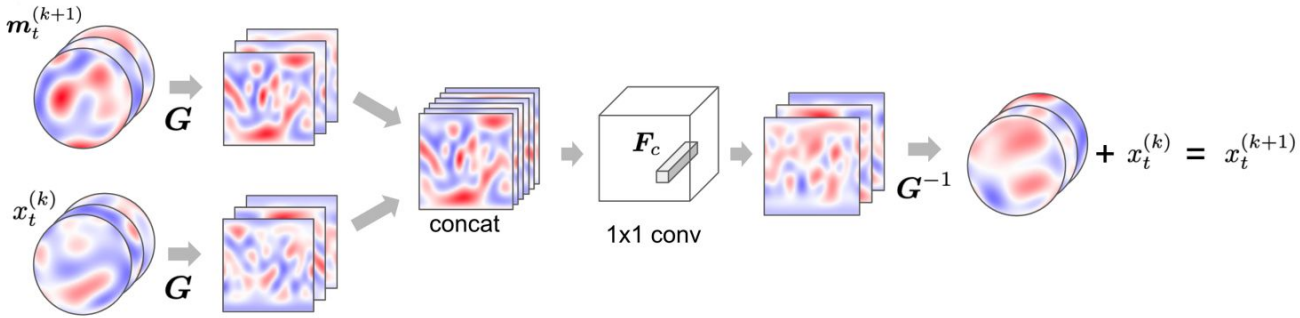
Spherical harmonics : Encode the orientation information of atoms in a molecule

Maintain rotational equivariance, 분자가 rotate해도 energy, force prediction을 동일하게 유지

Message passing with orientation : Atom embedding과 함께 edge orientation정보도 같이 전달

Message aggregation시에 edge message를 global coordinate frame으로 회전시켜 동일성 유지

Spherical harmonic representation과 spherical grid representation사이 변환, CNN 통해 embedding update



Competition results

		Energy MAE (eV)			
Rank	Team	Test-like	Rotated	Anomalous	Dense
1	TTRC (previously "Tencent AI Lab") Jiaqi Han[1], Tian Bian[2], Geyan Ye[3], Kaili Ma[2], Yuduo Zhi[3], Kangfei Zhao[3], Tingyang Xu[3], Wenbing Huang[4], Yu Rong[3] 1=Tsinghua University, 2=The Chinese University of Hong Kong, 3=Tencent AI Lab, 4=Renmin University of China	0.3960	0.3964	0.8904	0.3630
2	Atomic Architects MIT Yi-Lun Liao, Tess Smidt Massachusetts Institute of Technology	0.4266	0.4235	0.9228	0.3838
3	XJTUNRTeam	0.5255	0.5176	1.1577	0.4718
3	AutoGraph Xu Wang, Huan Zhao 4Paradigm	0.5263	0.5176	1.0789	0.4933
5	Shanghai Jiao Tong University Wei Yang[1], Frank Ji[1], Yulian He[2], Cheng Hua[3], Guanjie Zheng[3], Zhanyu Liu[3], Feixiang Tian[2], Tianhua Li[3], Junlin He[3] 1=Yalotein Biotech, 2=University of Michigan - Shanghai Jiao Tong University Joint Institute, 3=Shanghai Jiao Tong University	0.6529	0.6603	1.2163	0.5875
6	personal test Bangjian Zhou[1], Ji Wei Yoon[1], Zhuoyi Lin[1], J Senthilnath[1], Chaitanya K. Joshi[2] 1=Agency for Science, Technology and Research, Singapore, 2=University of Cambridge	1.0709	1.0671	1.5259	0.6058

2022 OC Challenge result

Rank	Team	Success Rate (%)
1	CausalAI Xinyu Li[1,2], Zhen Zhang[1,2], Anton van den Hengel[1,2], Javen Qinfeng Shi[1,2] 1=The University of Adelaide, 2=Australian Institute of Machine Learning	46.0
2	The Italian Job Luigi Bonati, Simone Perego, Pedro Buigues, Pietro Novelli, Riccardo Grazzi, Massimiliano Pontil Italian Institute of Technology	39.5
3	Robo Space	38.5
3	UCB ASK	22.5
5	LLM-JAR Janghoon Ock[1], Rishikesh Magar[2], Akshay Antony[3], Amir Barati Farimani[1] 1=Carnegie Mellon University, 2=Deep Apple Therapeutics, 3=Arcbest Technologies	5.0
6	AI Eng	0.50

2023 OC Challenge result

해 볼 만한 연구인지: 연구의 설득력, 백그라운드

결과물의 영향력 낮음: Contest에서 가장 성공한 모델의 영향력도 크지 않음. 문제상황이 확실하고 해결 방안의 임팩트가 클 것이라 생각했는데 생각과 다른 것인지

연구 진행 approach

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.)

AdsorbML: A Leap in Efficiency for Adsorption Energy Calculations using Generalizable Machine Learning Potentials (Lan et al.)

Spherical Channels for Modeling Atomic Interactions (Zitnick et al.)

Reducing SO(3) Convolutions to SO(2) for Efficient Equivariant GNNs (Zitnick et al.)