Reimagining Catalyst Reactivity Prediction with Machine Learning: From Data to Software Development

Xue Zong

Advisor: Dionisios G. Vlachos

Committee Members: Stavros Caratzoulas, Raul Lobo, & Antony Beris

---

Heterogeneous catalysis is crucial in various industrial processes impacting energy conversion, chemical manufacturing, and environmental remediation. Advances in catalyst design could mitigate climate change and create a sustainable future. However, catalyst design is hard to achieve experimentally due to the complexity of catalyst structures under working conditions. First-principles-based multiscale modeling simulates chemical reactions across scales and elucidates complex chemical systems. This thesis develops structure-dependent multiscale modeling frameworks assisted by machine learning (ML) for heterogeneous catalytic reactions. ML can enhance prediction accuracy and accelerate material screening.

First, we combine density functional theory (DFT) calculations and ML to create two models for accurately predicting the adsorption of small species on transition metal surfaces. The first model employs the traditional linear scaling relation using the generalized coordination number (GCN) as a descriptor. Our findings indicate that while the GCN scaling relations are effective for oxygen and high-valency carbon species, they are less accurate for others. To overcome these limitations, we develop a minimalistic ML model with easily computable properties of gas-phase species and adsorption sites. The generality of our methodology is validated by training the model using literature data on transition metals and successfully applying it to predict adsorption energies on single-atom alloys. These models could predict the structure sensitivity of surface reactions and improve catalytic performance via particle engineering.

Second, we introduce a methodology for constructing a structure-descriptor-based microkinetic model to explore nanoparticle structure-reaction kinetics relations. We illustrate this approach using the complete oxidation of methane on Pt as a model reaction. A volcano-like rate

is observed with an optimum coordination number; small nanoparticles have low reactivity because of carbon poisoning. This methodology enables rapid prediction of kinetic performance and active site determination for designing optimal catalyst structures. To streamline the workflow of model construction, we develop an open-source Python software package named Descriptor-based Microkinetic Analysis Package (DescMAP), designed to automate descriptor selection, volcano curve generation, and active site identification for structure-sensitive reactions.

Finally, we develop models to estimate activation energies for hydrogenolysis reactions of large hydrocarbons on metal surfaces. Traditional Brønsted -Evans-Polanyi (BEP) relationships predict dehydrogenation reactions more accurately than C-C cracking reactions. The correlations are more accurate on Pt compared to Ru and Rh. However, BEP estimations result in relatively significant errors when consolidating all data. ML models are established to reduce prediction error. We also assess the model's transferability in predicting activation barriers across various transition metals. These advancements can potentially expedite catalyst screening, marking a valuable contribution to heterogeneous catalysis.