# Machine Learning for Heterogeneous Catalyst Design and Discovery

**Takashi Toyao**
Assistant Professor
Affiliation: Institute for Catalysis, Hokkaido University
e-mail: toyao@cat.hokudai.ac.jp
Specialty: Heterogeneous catalysis

**ABSTRACT**

Molecular/materials informatics has become a central paradigm in molecular and materials science thanks to the enormous potential it holds to revolutionize the design of functional molecules/materials. However, although we have already seen proof-of-concept examples that artificial intelligence (AI) can reduce the time and costs involved and also can find new compounds, most of them have only been tested on benchmark problems and no fundamentally-new molecules/materials or synthetic-transformations have been found. This is primarily due to lack of data and that machine learning (ML), the main player in this campaign, is highly focused on optimization rather than finding novel compounds and phenomena (extrapolation).

Establishing "Catalysis Informatics" is even more challenging.[1] Although it is highly related to materials informatics and chemoinformatics, it is distinguished by the fact that catalysis is a time-dependent dynamic event controlled by the structures and chemical nature of catalytically active sites. In particular, heterogeneous catalysis is still a largely empirical science due to the complexity of the surface chemistry involved. This situation causes lack of data as the computational costs to obtain accurate theoretical models for such complex heterogeneous catalysis are currently prohibitively high and high-throughput experimental methods, which have been applied successfully to relevant fields, have not been explored fully at the current time. In this regard, building ML models that effectively find novel catalysts within diverse chemical space from "real world" experimental catalysis data (not from well-behaved computational data) is highly desirable.

In this context, our group has made a new ML approach which uses elemental features as the input representations rather than inputting the catalyst compositions directly.[2,3] Namely, in our proposed method, the elemental composition ratios are multiplied by elemental descriptors such as electronegativities, melting points, atomic radii, etc. which are unique for each element. For demonstration of our ML approach, literature data based on the oxidative coupling of methane (OCM) and water gas shift (WGS) reactions have been analyzed. Our ML approach, which considers elemental features as input representations rather than the catalyst compositions, was successfully applied, and new promising catalyst candidates for future research were proposed. Latest results on $CO_2$ hydrogenation will also be shown in the presentation.

**REF.**

[1] T. Toyao, Z. Maeno, S. Takakusagi, T. Kamachi, I. Takigawa, K. Shimizu, Machine Learning for Catalysis Informatics: Recent Applications and Prospects, ACS Catal. 10 (2020) 2260–2297.

[2] K. Suzuki, T. Toyao, Z. Maeno, S. Takakusagi, K. Shimizu, I. Takigawa, Statistical Analysis and Discovery of Heterogeneous Catalysts Based on Machine Learning from Diverse Published Data, ChemCatChem. 11 (2019) 4537–4547.

[3] S. Mine, M. Takao, T. Yamaguchi, T. Toyao, Z. Maeno, S.M.A. Hakim Siddiki, S. Takakusagi, K. Shimizu, I. Takigawa, Analysis of Updated Literature Data up to 2019 on the Oxidative Coupling of Methane Using an Extrapolative Machine-Learning Method to Identify Novel Catalysts, ChemCatChem. 13 (2021) 3636–3655.