Prediction of reaction yield for high throughput experimental data sets by deep learning

Akinori Sato, Nara Institute of Science and Technology

Chemical reaction yield is an important factor to determine reaction conditions. Predicting yield for reactions which have never been done is difficult. Highly accurate yield prediction models could be used for the rapid determination of the optimal reaction condition. Our final goal is to build a highly predictive yield prediction scheme applicable to diverse reaction classes. As a first step to the goal, yield prediction for homogenous high-throughput experimentation (HTE) data was investigated. Employed HTE data sets were for Buchwald-Hartwig cross coupling and Suzuki-Miyaura cross coupling reactions, which are commonly used for testing predictive performances of machine learning-based yield prediction models. We have developed several neural network architectures, such as using Mol2Vec as atom embedding features in molecular graphs, molecular contrastive learning as pre-training and incorporating Transformer for diverse chemical reactions. Overall, our model performances are comparable with those of other state-of-the-art models. In this presentation, I will present our approach to data-driven yield prediction, current situations and future work.