Multicomponent Reaction Optimization

for Selectivity in Asymmetric Catalysis

Philippe Gantzer[1], Vitor Alcantara Fernandes[1], Haruto Yamato[1], Tynchtyk Amatov[2], Nobuya Tsuji[1], Benjamin List[1,2], Pavel Sidorov[1]

[1]*Institute for Chemical Reaction Design and Discovery (WPI-ICReDD), Hokkaido University, Sapporo 001-0021, Japan.*

[2]*Max-Planck-Institut für Kohlenforschung, 45470 Mülheim an der Ruhr, Germany*

Optimizing multicomponent reactions in regards of the composition of reaction mixture and reaction conditions is challenging due to the vast chemical space generated by the combinatorial variation of reactants and the high cost associated with experimental screening. We present here an automated, iterative machine learning-based workflow designed to reduce the number of experiments required to identify optimal reaction conditions.

The proposed process begins by clustering possible combinations and selecting an initial set of reactions to perform, in order to broadly sample the chemical space while minimizing synthetic effort. A Gaussian Process regression model is then built on this initial dataset, using chemical structures encoded by fragment-count based molecular descriptors such as the ones available in DOPtools library.[1] The outcome of the model (predicted values and their uncertainty) is then used in a Bayesian Optimization approach to guide the selection of subsequent candidate reactions based on prior results. An acquisition function balances exploration of uncertain regions and exploitation of promising ones, and the loop is iteratively repeated with new data until satisfactory results are obtained.

In this presentation, we will discuss several applications of this workflow. First, the benchmark on the dataset of Santanilla *et al.*,[2] comprising an exhaustive exploration of 1,536 palladium-catalyzed C–N cross-coupling reactions with various nucleophiles, bases, and catalysts, has the objective of identifying the highest-yielding conditions in as few steps as possible. In this benchmark, we test various Gaussian Process model hyperparameters such as kernels and number of iterations, as well as the influence of the composition of the initial training set and batch size on the outcome of the Bayesian Optimization. After 38 initial experiments, our approach required as few as 16.5 median additional experiments to reach the objective.

Furthermore, this workflow is currently employed in a collaborative project dedicated to the iterative selectivity optimization in asymmetric catalysis with large organic catalysts. The key challenge here lies in balancing the need for substrate specificity with catalyst generality, as the resulting enantiomeric ratio and the reactivity of the mixture depend heavily on the structural composition of all components.

Bibliography :

[1] Byadi et al., Digital Discovery 4 (2025) 1188-1198.
[2] Santanilla et al., Science 347 (2015) 347, 49–53.
[3] A. Mushtaqa, A. F. Zahoor, RSC Adv. 13 (2023) 32975.