Multicomponent Reaction Optimization with Gaussian Processes on Fragment Count Descriptors

Philippe Gantzer
Hokkaido University

Optimizing multicomponent reactions is challenging due to the vast chemical space generated by the combinatorial variation of reactants and the high experimental cost associated with screening. We propose an automated, iterative workflow designed to reduce the number of experiments required to identify optimal reaction conditions. The process begins by clustering possible combinations and selecting an initial set of reactions that broadly sample the chemical space while minimizing synthetic effort. A Gaussian Process model then guides the selection of subsequent candidates based on prior results using Bayesian optimization. Unlike previously reported methods, our approach employs fragment-count based molecular descriptors, which produce a larger feature space but better capture structural diversity. We validated this workflow using a literature dataset, successfully identifying the highest-yielding conditions among all possible combinations. Future work will extend this framework to asymmetric catalysis, where a key challenge lies in balancing the need for substrate specificity with catalyst generality, a trade-off that often limits substrate scope.