Rapid evaluation of virtually synthesized compounds using a support vector machine model with reactant-wise kernels

Yuto Iwasaki
Nara Institute of Science and Technology

We propose a support vector regression (SVR)-based virtual screening workflow to rapidly evaluate reactant combinations to propose virtual compounds with high predicted activity. A key aspect of our approach is to use reactant-wise kernel functions instead of a kernel function to evaluate a pair of molecules in SVR. Since a tuple of reactants to form a product might lose the information of a whole molecular structure, a data augmentation technique was introduced by virtually dissecting product compounds to generate reactant tuples, which are included in the training dataset. A benchmark calculation across 60 reaction datasets targeting 10 macromolecules showed that the proposed SVR models achieved comparable prediction accuracy to standard SVR models. As a demonstration, 6.4 trillion reactant pairs were exhaustively evaluated within 8 days on a single CPU machine, confirming the method's scalability.