Novel Drug Candidate Design and Retrosynthetic Translation using Artificial Intelligence

Lee, Juyong

Department of Chemistry, Division of Chemistry and Biochemistry, Kangwon National University, Chuncheon, Republic of Korea
E-mail: juyong.lee@kangwon.ac.kr

Here, we present a new molecule optimization method, MolFinder, based on an efficient global optimization algorithm, the conformational space annealing algorithm, and the SMILES representation. MolFinder finds diverse molecules with desired properties efficiently without any training and a large molecular database. Compared with recently proposed reinforcement-learning-based molecule optimization algorithms, MolFinder consistently outper- forms in terms of both the optimization of a given target property and the generation of a set of diverse and novel molecules. The efficiency of MolFinder demonstrates that combinatorial optimization using the SMILES representa- tion is a promising approach for molecule optimization, which has not been well investigated despite its simplicity. We believe that our results shed light on new possibilities for advances in molecule optimization methods. In addition, we also present a new retrosynthesis prediction method using substructural-based tokenization combined with a transformer architecture. We captured the chemical change by using the atomic environments as fragmentation patterns associated with the bits of Morgan fingerprints. Describing a molecule with a complete set of atom environments establishes a clear relationship between translated product-reactant pairs due to conservation of atoms in reactions. We investigated the effect of different encoding scenarios on predicting the reactant candidates and critically assessed the retrieval process with respect to coverage, degeneracy and resolution. Our model achieved a top-1 accuracy of 66.5% within the bio-actively similar range for USPTO test dataset, outperforming the other state of the art, translation methods. The end-to-end, template-free model for retrosynthetic prediction provides fast and reliable retrosynthetic route planning for substances whose fragmentation patterns are revealed.