A simple, but powerful machine learning method for efficient processing of chemical structures

진서안, 김현욱†
Department of Chemical and Biomolecular Engineering (BK21 Plus Program), Korea Advanced Institute of Science and Technology (KAIST)
(ehukim@kaist.ac.kr†)

When it comes to material design or drug design, there exist numerous chemical candidates to consider. Importantly, experimental validation for such a large volume of chemicals is prohibitive in terms of time and efforts. In addition, side effect and toxicity should also be taken into account when designing drugs. Here, a simple, but powerful machine learning method, named dendrite perceptron (DP), is proposed to predict chemical properties and relevant substructures of a chemical structure given as an input. In contrast to a deep neural network having multiple layers, DP has a single layer architecture, but generates more precise predictions for several intended applications. DP was examined on three different datasets from DrugBank, Tox21 database and dye-sensitive solar cell database (DSSCDB). As a result, the DP shows higher accuracies and predictions than existing state-of-the-art methods, and further suggested substructures that are important for the observed chemical properties. The DP is expected to help our better understanding and more efficient processing of a large number of chemical structures that are important in chemical and biological systems.