Prediction of Molecular Properties with Graph-based Machine Learning Methods

Hyun Woo Kim

Department of Chemistry, Gwangju Institute of Science and Technology (GIST), Gwangju 61005, Korea

Recently, artificial Intelligence (AI), which includes machine learning (ML), is applied to the research of organic chemistry. Predicting properties or retrosynthesis of organic molecules is an active research topic. To this end, organic molecules should be represented in a machine-readable format, which serves as an input variable to the ML algorithm. In many cases, these representations are based on physics and chemistry. Here, we present a data-driven method to improve molecular representations. As this method is based on a graph neural network (GNN), I will briefly describe an example of a GNN called graph convolutional network (GNN). The working code of GCN will also be introduced. Then I will describe how to extend deep metric learning to find data-driven representations. We tested new representations on several molecular datasets and found that the ML algorithm using the new method outperformed the old one. More importantly, new representations are powerful when the dataset size is small.