Multi-property prediction model based on patents and PoLyInfo database

A. Nakao, CrowdChem

With the development of deep learning, machine learning based on large data sets have attracted attention in many fields. Patent data, which is a large and open dataset of technical texts, is a potentially promising data sources for utilizing machine learning techniques in the field of materials and product development. We constructed a machine learning model through multi-task learning based on the wide variety of experimental data described in patents. The model can be adapted to a wide range of applications and used as a web application. However, patents only describe experiments enough to identify the invention, and the quality of the data is not sufficient to make accurate predictions. This study aims to improve the accuracy of patent-based property prediction models by performing multi-task learning with high-quality data in PoLyInfo, a large polymer database constructed by the National Institute for Materials Science.