Mining Insights on Metal-Organic Framework Synthesis from Scientific Literature Texts

강영훈, 박현수, 김지한[†]
KAIST
(jihan929@gmail.com[†])

Identifying optimal synthesis conditions for metal-organic frameworks (MOFs) is a major challenge that can serve as a bottleneck for new materials discovery and development. Trial-and-error approach that relies on a chemist's intuition and knowledge has limitations in efficiency due to the large MOF synthesis space. To this end, 47,187 number of MOF were data mined using our in-house developed code to extract their synthesis information in papers. The text-mining algorithm yields an average F1 score of 90.3 % across different synthesis parameters. From this data set, a PU learning algorithm was developed to predict synthesis of a given MOF material using synthesis conditions as inputs, and this algorithm successfully predicted successful synthesis in 83.1 % of the synthesized data in the test set. Finally, our model correctly predicted three amorphous MOFs as having low synthesizability scores while the counterpart crystalline MOFs showed high synthesizability scores. Our results show that big data extracted from the texts of MOF papers can be used to rationally predict synthesis conditions for these materials, which can accelerate the speed in which new MOFs are synthesized.