PB 11

Prediction of Melting Point and Fusion Enthalpy of Cocrystal by Machine Learning combined with molecular informatics

Authors and affiliation

Yingquan HAO[1] and Yusuke Shimoyama*[,1]

[1]Department of Chemical Science and Engineering, Tokyo Institute of Technology, Japan

*E-mail: yshimo@chemeng.titech.ac.jp

Key Word (3 words)

Cocrystal, Machine Learning, Chemical Calculation

Abstract (less than 300 words)

In the past decades, the solubility of the pharmaceuticals become lower. To solve this problem, the cocrystal have been considerate. By forming a new crystal structure with an additive which is called as coformer (CF), the thermodynamics property like melting point, fusion enthalpy of the pharmaceutical crystal can be modified.

But by now, the melting point, fusion enthalpy of the cocrystal is mostly collected by experiments. So, in this research, we will try to build a machine learning model which can predict melting point and fusion enthalpy of cocrystal only from the molecular information of pharmaceutical (API) and coformer.

To make the molecular understandable for computer, we use the statistical charge distribution calculated by quantum calculation based on Conductor-like Screening Model (COSMO) as the descriptor for API and coformer. In additional, because the fusion enthalpy and melting is strongly related with energy, the sigma-profile of the API and coformer is input as a heat map matrix. The routable bond number, fusion enthalpy and melting point of the API and coformer are also inputted. For machine learning, least absolute shrinkage and selection operator with cross validation (LassoCV), support vector regression model with linear kernel (SVR-linear) is applied.

About 55 data sets are collect from the published paper. As the final result, the R^2 score, the mean relative error (MRE), and mean absolute error (MAE) for each test datasets are as below. For the melting point prediction, LassoCV model gives out the best result, while R^2 = 0.669, and the MRE =3.9 %, MAE=16 K. And for fusion enthalpy prediction, SVR-linear model shows best performance, give out a R^2 = 0.625, MRE =12.4% and the MAE=8.5 kJ mol^{-1}. For both cases, the R2 is higher than 0.6 which indicating strong correlation between prediction and experiment, and the error is acceptable.

MTMS ’21