PB 06

Calculation of Solubility of Organic Compounds in Supercritical Carbon Dioxide Using Machine Learning with Molecular Descriptors

Authors and affiliation

Naoya Maeda[1], Junichi Sakabe[2], Hirohisa Uchida[*2]
[1] Graduate School of Natural Science and Technology, Division of Natural System, Kanazawa University
[2] Faculty of Frontier Engineering, Institute of Science and Engineering, Kanazawa University
*E-mail: uchida@se.kanazawa-u.ac.jp

Key Word (3 words)

Solubility, Machine learning, Molecular descriptors

Abstract (less than 300 words)

The solubility of target substances in supercritical CO_2 ($scCO_2$) is essential as a basic knowledge for designing the processes and finding the optimum conditions of industry processes [1]. The solubility can generally be determined by experimental measurements or calculations. However, experimental measurements are very time-consuming and costly. In addition, the thermodynamic calculation methods, which are widely used nowadays, require various physical properties and interaction parameters of solutes and solvents, leading it difficult to calculate the solubility easily.

In this study, the solubility of 29 organic compounds in $scCO_2$ was calculated using neural network based on the Quantitative Structure-Property Relationship (QSPR) model [2]. We used 886 solubilities of 29 organic compounds in $scCO_2$ under various temperatures and pressures. The solubility data of organic compounds were randomly divided into training, validation and test sets. The training data and the validation data each contain data of the same compound under different conditions. Test data is new data that was not used for training and evaluation data. A large number of molecular descriptors were calculated using RDKit [3]. The descriptors for all compounds with the same or zero data values and those with pair correlation greater than 0.9 were eliminated. In the calculations, we used four parameters such as temperatures, pressures, 99 descriptors for the organic compounds, and the density of CO_2. Modeling of the relationship between the parameters and solubility data was achieved by artificial neural network method. The artificial neural network architectures and hyper-parameters were optimized empirically.

The calculation results showed that the solubility of many compounds could be expressed qualitatively and quantitatively relatively well. However, there were some compounds whose solubility was difficult to express quantitatively. These may be improved by considering the interactions between the compounds and CO_2.

References

[1] A. J. A. Meirelles, Fluid Phase Equilib. 482 (2019) 1-10.

[2] R. Tabaraki, A. Toulabi, Fluid Phase Equilib. 383 (2014) 108-114.

[3] https://www.rdkit.org/.