Towards Transferable User-Friendly Machine Learning for Thermophysical Property Prediction

Frank Mtetwa[1, S], Thomas Knotts[1, C], Vincent Wilding[1] and Neil Giles[1]

[1]*Chemical Engineering, Brigham Young University, Provo, UT, U.S.A.*
thomas.knotts@byu.edu

In the rapidly advancing landscape of diverse fields and industries, the integration of machine learning (ML) to enhance processes has become a pervasive pursuit. The application of ML in predicting thermophysical properties has witnessed a surge in interest, leading to a proliferation of research papers on the subject. Despite this enthusiasm, widespread adoption of ML models faces challenges, primarily stemming from the inherent "black box" nature of the technique. Reproducing the work of others is hindered by issues such as the absence of provided datasets, incomplete specification of features, and the withholding of the model itself in literature.

This talk addresses these challenges, shedding light on their impact on model performance. Using the normal boiling point and critical point of a compound as a case study, the model is trained, validated, and tested using data from the DIPPR 801 database. Notably, the focus is on exploring various approaches to feature selection and their consequential effects on model performance. The talk also delves into the complexities of developing ML approaches for chemical properties, emphasizing the need for an optimal set of features and the challenges associated with selecting descriptors. A significant advancement in this work involves the incorporation of transfer learning. Leveraging a pretrained model for normal boiling point prediction, the presentation demonstrates how transfer learning can be employed to predict critical temperatures. This approach not only enhances efficiency but also showcases the potential for leveraging existing knowledge to tackle new challenges.

The talk concludes by unveiling an optimal model designed for accurate property prediction, distinguishing itself through user-friendly features for external utilization. The model simplifies the input process by requiring only the SMILES notation of a compound—an easily obtainable piece of information. Developed with Pytorch, the model's transferability is facilitated through a trained model, enabling other research groups to seamlessly apply the technique to predict normal boiling points for their compounds of interest. This work not only advances the understanding of ML in predicting thermophysical properties but also promotes accessibility and collaboration within the scientific community.