FOPAM 2026
Proceedings of the 3rd Foundations of Process/Product Analytics and Machine Learning
Decatur, USA, July 26-30 (2026)

Optimal Solvent Mixture Screening with Graph Neural Networks

Yipei Zhao

University of Sheffield, School of Chemical, Materials and Biological Engineering

ABSTRACT

Solubility is a critical parameter governing the efficacy and stability of agrochemical active ingredients (AIs) formulated in solvent mixtures. Traditionally, the screening of efficient solvent mixtures heavily relies on empirical, trial-and-error experiments. The nonlinear behaviours of complex solutes in multi-component solvent systems remain difficult to predict. Furthermore, while predictive machine learning models have shown success in predicting the interaction between multi-component solvents, the application of multi-component system solubility prediction is lacking in the agrochemical field. To address this gap, our study employs a Graph Neural Network (GNN) to model the behaviour of dissolving systems from both intramolecular and intermolecular perspectives. Rather than altering the underlying structural architecture, this work focuses on the domain-specific application and transferability of the model to a novel, high-standard experimental dataset of agrochemical AIs and targeted solvent mixtures. By training the model exclusively on specialised agrochemical data, we bypass the limitations of generalised chemical databases and improve the model's predictive accuracy for our specific use case. This work highlights the critical importance of agrochemical-specific data application in applied machine learning for chemical engineering. Ultimately, we demonstrate how leveraging established GNN architectures alongside specialised datasets can provide a highly scalable, pre-screening tool that can drastically reduce the cost of experiments in the development of novel agrochemical formulations.