Completing the Matrix for Efficient Process Design – Advancing the Prediction of Thermodynamic Parameters and Physical Properties by Machine Learning

Tobias Averbeck,[1] Alexander Kühl,[1] Gabriele Sadowski,[2] Christoph Held,[2] Thomas Seidensticker[2]
[1] Laboratory for Industrial Chemistry, TU Dortmund University
[2] Laboratory of Thermodynamics, TU Dortmund University

Abstract

In the field of chemical engineering and thermodynamics, the accurate prediction of activity coefficients (γ values) plays a pivotal role in modeling and understanding the behavior of mixtures concerning their phase equilibria. These γ values, crucial for designing efficient separation processes and reaction media, offer insights into non-ideal interactions within mixtures. Traditional approaches for determining γ values, including experimental measurements and thermodynamic models, face significant challenges. Experimental methods are notably time-intensive and resource-demanding, often limited by the tedious equilibration required for accurate vapor-liquid equilibrium assessments. Consequently, pursuing more efficient, predictive models has led to exploring various computational strategies.

Recent advances have underscored the effectiveness of group-contribution methods, quantum chemistry, and machine learning (ML) in forecasting γ values.

Equations of state like PC-SAFT require binary interaction parameters, which are typically adjusted or determined using group contribution methods. Moreover, it is feasible to adjust these parameters based on properties predicted through machine learning techniques. Although existing tensor completion methods have made progress in predicting two- and three-dimensional data, they still struggle to effectively capture nonlinearities and temporal dependencies in relational data.

We introduce the novel 3D-DMF-H method for tensor completion,[1] a neural network-based approach extending Deep Matrix Factorization (DMF) to handle nonlinear data structures and incorporate additional data points. This method is versatile for various three-dimensional tensor completion problems and achieves exceptional accuracy in predicting activity coefficients for phase equilibria modeling. Remarkably, 3D-DMF-H surpasses the UNIFAC thermodynamic model and improves azeotrope predictions with PC-SAFT. For instance, from about 110,000 data points available in the Dortmund Database (DDB), 7.5 million new data points were predicted. Compared to UNIFAC, our method has a five-times lower mean absolute error of around 0.11. which is within the measurement uncertainty for determining activity coefficients empirically.

These results underscore the potential of machine learning in the chemical industry and the need for ongoing algorithmic refinement and exploration.

We are now exploring the 3D-DMF-H method's applicability in new areas like solubility prediction and membrane flows, offering a nuanced and comprehensive tool that bridges the gap between theoretical models and practical applications.

[1] T.Averbeck, G. Sadwoski, C. Held, T. Seidensticker, *Ind. Eng. Chem. Res.*, **2024**, accepted.