Prediction of Activity Coefficients with Matrix Completion Methods

Fabian Jirasek [C, S] and Hans Hasse
Laboratory of Engineering Thermodynamics (LTD), TU Kaiserslautern, Kaiserslautern, Rhineland-Palatinate, Germany
fabian.jirasek@mv.uni-kl.de

Methods for predicting fluid properties of mixtures are of fundamental importance in process engineering and highly developed physical methods are available, such as UNIFAC or COSMO-RS. In the present work, an alternative approach to predict fluid properties of mixtures based on machine learning is presented. It uses the fact that data on binary mixtures can conveniently be stored in matrices. When applied to thermodynamic properties, these matrices are usually only sparsely occupied, as experimental data are available for only a small fraction of the potentially interesting mixtures. The prediction of the unobserved entries constitutes a matrix completion problem, which is well-known in machine learning. Different methods for its solution are available, whose development has been triggered by the *Netflix Prize* [1], an open competition that aimed at improving the recommender system of Netflix. In this work, we demonstrate that these matrix completion methods (MCMs) can also be applied successfully to thermodynamic problems [2,3]. As an example, activity coefficients are considered here, but the approach can be applied to other fluid properties straightforwardly. In our MCM, the algorithm learns latent features of the two components of a binary mixture from the available data on activity coefficients using Bayesian inference. The learned features then enable predictions for the activity coefficients in all mixtures of the considered components, including mixtures for which no data are available. This purely data-driven MCM gives excellent predictions, even though no physical descriptors of the studied components are used. In an example of its application, we show that it outperforms the current benchmark modified UNIFAC (Dortmund) in predicting activity coefficients at infinite dilution. Furthermore, ways to combine the data-driven MCMs with physical thermodynamic models were studied, which lead to further improvements of their performance. In contrast to physical models, the MCM-based mixture models can be retrained easily when new data become available or only subsets of data are of interest. MCMs open the way to a new generation of methods for predicting thermodynamic properties of mixtures.

References

[1] J. Bennett, S. Lanning, The netflix prize. *Proceedings of KDD cup and workshop* (2007) 35.

[2] F. Jirasek et *al., J. Phys. Chem. Lett.* 11 (2020) 981-985. DOI: 10.1021/acs.jpclett.9b03657.

[3] F. Jirasek, R. Bamler, S. Mandt, *Chem. Commun.* 56 (2020) 12407-12410. DOI: 10.1039/D0CC05258B.