Improved Activity Predictions by Transfer Learning the COSMO-SAC Model

Eric Fonseca

Department of Materials Science & Engineering, University of Florida, Gainesville, FL

Thermodynamic activity models are critical to chemical process design and to predict physical properties of pure materials and mixtures. Using artificial neural networks and freely available databases of sigma profiles, we generated a model capable of reproducing the COSMO-SAC activity model outputs. We used 1060 molecules from the Virginia Tech Sigma Profile Database to generate activities for 540,000 unique binary mixtures at a fixed temperature of 300 K. Mimicking the COSMO-SAC thermodynamic activity model, our model's inputs are mole fraction, sigma profile, surface area, and volume. The models achieve a mean absolute error of 0.05 compared to COSMO-SAC. Across all compositions $\ln(\gamma_i^{\infty})$ the model achieves a 0.101 mean absolute error relative to COSMO-SAC. These pre-trained models are then used to transfer learn with experimental data gathered from the NIST ThermoML dataset. Our results show that our ML models greatly outperform naïve ML approaches and COSMO-SAC predictions by up to 40%. This approach also can aid in improving COSMO-SAC formulations by analyzing the resultant neural network architecture and parameters.