Machine-learning-powered molecular design: optimal solvents for hybrid extraction-distillation

Johanna Lindfield, Luca Bosetti, Benedikt Alexander Winter, Johannes Schilling, and André Bardow*
Energy and Process System Engineering, ETH Zurich, 8092 Zurich, Switzerland

Keywords: molecule design, process design, thermodynamic properties

Computer-aided molecular and process design (CAMPD) simultaneously optimizes molecules and processes, relying on accurate molecular property predictions. Group contribution (GC) and quantum mechanics-based (QM) methods have traditionally been used (Papadopoulos 2018), but machine learning (ML)-based frameworks have recently demonstrated higher accuracy and efficiency due to their ability to capture complex relationships. A natural language model has been shown to outperform GC and QM methods in calculating activity coefficients for various molecules (Winter 2022).

The full potential of incorporating ML-based property prediction into the CAMPD framework remains unexploited. The accuracy and speed of property predictions make this approach promising for exploring a larger design space in an efficient manner. In this work, we combine ML-based property prediction and genetic algorithm-based (GA) molecule design with pinch-based process models to optimize solvent molecules for extraction-distillation.

The present investigation focuses on the hybrid extraction-distillation process for the separation of γ-valerolactone (GVL) from aqueous solutions. GVL is a promising bio-based platform chemical. However, its separation is challenging. Our aim is therefore to reduce the energy demand for separation. While the conventional process involves only distillation, the inclusion of an extraction unit is expected to lower the overall energy consumption and production costs (Scheffczyk 2018). The liquid-liquid equilibrium between water and a suitable solvent is exploited in the extraction unit, which necessitates careful consideration of the solvent design.

Method and results

To include an extraction step in a process, a suitable solvent must be identified, which requires precise methods to calculate its thermodynamic properties. The SMILES-to-Properties Transformer (SPT) machine learning model, is a natural language processing model and takes SMILES code as input to predict molecular properties. SPT is originally trained to estimate activity coefficients (Winter 2022). Here, we further trained SPT to estimate properties relevant to extraction processes, such as melting and boiling temperatures. The thermodynamic properties can either be used directly as heuristic objectives for solvent design by the GA or serve as input for the process model to calculate a process-level objective function, such as the energy demand for the entire hybrid extraction-distillation process. The GA iteratively ranks and refines the solvent design according to the objective function.

In this abstract, we only report the solvent design based on the minimum energy demand of the entire process, which is driven by the heat required in the reboiler of the distillation column. By generating solvents composed of only carbon, oxygen, and hydrogen atoms, the best separation performances, according to the minimum energy demand, have been achieved with isobutyric acid ($Q_{min} = 3.814 \text{ MJ/kgmole}_{\text{feed}}$) at a GA runtime of 8h. When comparing isobutyric acid to the best-performing molecule of Scheffczyk et al. (2018), divinyl ether, it was found that using isobutyric acid resulted in a 14.8% reduction in the minimum energy required for the hybrid extraction-distillation process.

Conclusion

In summary, we introduce a machine-learning-powered framework for the solvent design of the separation of a GVL from water. This design framework allows us to screen rapidly and accurately a vast number of solvents for which the thermodynamic properties are not experimentally available. This framework can be expanded to include conceptual design, enabling the integrated computer-aided molecular and process design (CAMPD).

Acknowledgments

The project leading to this application has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 101036854 and from NCCR Catalysis (grant number 180544) a National Centre of Competence in Research funded by the Swiss National Science Foundation.

References


* To whom all correspondence should be addressed: abardow@ethz.ch