Workflow for Building and Analyzing Machine Learning Models based on Rigorous Flowsheet Simulations

Jan Schöneberger[1,2], Burcu Aker[1,2], Armin Fricke[1,2],

[1]Chemstations Europe GmbH, [2]Capital-Gain Consultants, Berlin, Deutschland

Flowsheet simulators solve the material balances, equilibrium equations, sum equations, and energy balances (MESH equations) for connected unit operations. Thermodynamic models for the phase equilibrium, the reaction equilibrium, and the energy balance introduce high non-linearities. This makes rigorous flowsheet simulations hard to solve, up to a level where no solution exists to the specified problem. Such simulations are quite expensive by means of computational effort. Close to these infeasible regions, the calculation time increases even further, reaching its maximum inside the infeasible region.

Machine learning (ML) models are explicit feed forward models that can be evaluated very fast. Therefore, it is a promising approach to replace the time-consuming solution of the MESH equations with ML equations.

This work addresses the workflow of building and validating ML models based on the example of the rigorous simulation of a prereformer reactor. The problem has 3 output variables and 11 design variables but only 10 degrees of freedom, because the reactor inlet mole fractions are correlated by the sum equation.

In a first step, an experimental plan is created by latin hypercube sampling (LHS). Then, the computer experiments are performed in parallel on a cluster of process simulator instances. Third, the results are analyzed with common tools to find correlations and linear dependencies. This allows a reduction of the dimensions, either when simple dependencies or when insignificant design variables are detected. In the next step, a ML model is selected and trained. This model is used to generate more data points in the 11D space. The low evaluation time of the ML model allows the evaluation of factorial designs on a fine grid. Iso-lines (lines with one variable changing while all others are kept constant) are generated and analyzed based on an engineer's understanding of the underlying physical process. Finally, the ML model is transferred back into the flowsheet simulator as a custom unit operation. There, its results can be directly compared with the rigorous calculations.

Once the ML model quality is satisfactory, it can be used for optimization, data visualization, or other purposes where many problem evaluations are required.