Machine learning assisted prediction of equilibrium points

In the process industry the use of process simulations is a widespread tool for many applications like process synthesis and optimization. A key part in simulations of most chemical processes is the description of thermodynamic phase equilibria in unit operations like distillation, extraction and decanting.
Modern models, like the Perturbed Chain-Statistical Associating Theory (PC-SAFT) [1] equation of state, are used for such purposes. An issue that arises from the usage of PC-SAFT is the computational complexity that results from an implicit equation that needs to be solved in a separate loop. This computational effort makes the applicability to process optimization and advanced process control methods limited.
Machine learning models provide a computationally cheap alternative to describe the underlying phase equilibria. These black-box models can be adapted to approximate any given functional relation. Because of their explicit structure they can be evaluated cheaply, which results in the general feasibility of process optimization, see Figure 1.

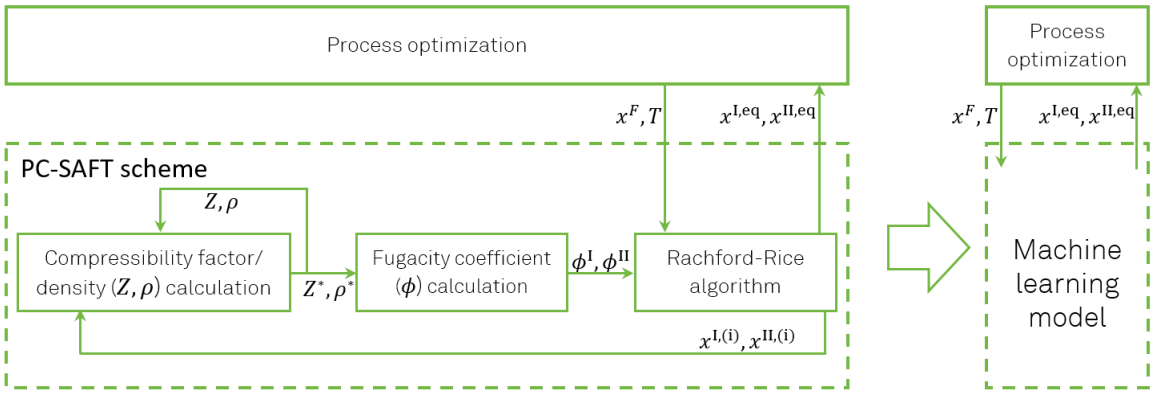
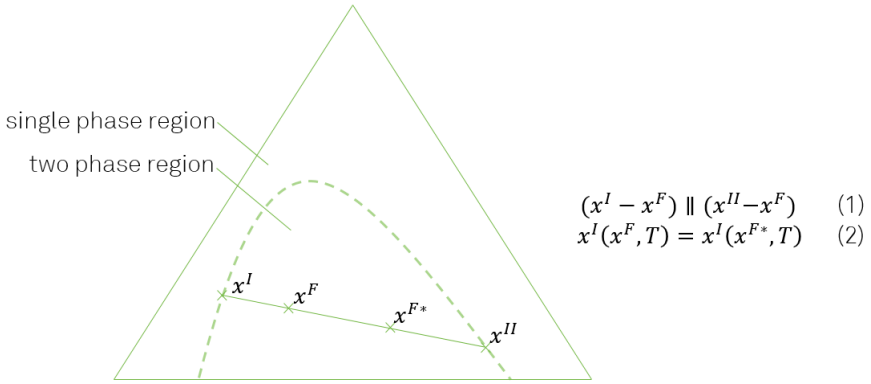


Figure 1: Simplification due to introduction of machine learning model

The description of thermodynamic phase equilibria is a challenging task. Ternary phase diagrams are used to represent all possible phases of three components and their equilibrium according to the composition of a mixture of the three components at constant temperature and pressure (Figure 2). The area enclosed by the dashed line is the two-phase region. The region that is not enclosed by the dashed line is the single-phase region. A point on the diagram (e.g. x_F) represents a composition that is specified in terms of mole or weight fraction. The response surface of the phase concentrations, while being smooth, shows significant nonlinearities. Additionally, the phase-compositions are only defined where multiple phases exist. In a liquid-liquid system this region is called the miscibility gap.

Task description

Your task is to develop a machine learning method to accurately predict the phase compositions as a function of feed composition and temperature. You are given a data set that was obtained by applying the Rachford-Rice algorithm to randomly initialized input variables. As in real life, the data firstly needs to be cleaned to determine the points which should be considered to be in the single-phase region versus those in the two-phase region.
For your model, you can consider the following hints/ ideas. In reality the two-phase compositions lie on a line with the feed composition (eq. 1). Given a certain temperature, every feed composition between two phase compositions results in the same phase compositions (eq. 2).



$$(x^I - x^F) \parallel (x^{II} - x^F) \quad (1)$$
$$x^I(x^F, T) = x^I(x^{F*}, T) \quad (2)$$

Figure 2: Visualization of equilibrium in a three component system

Mandatory tasks

The following tasks **have to be completed** in order to pass the project.

- Implement a machine learning model to approximate the phase compositions x^I, x^{II} as a function of feed composition x^F and temperature T. Consider that not every feed concentration will lead to phase separation.
- Investigate the model performance (in relation to the number of components and training set size).
- Compare different machine learning methods.

Note: Your tasks do **not** incorporate working with the PC-SAFT model or the Rachford-Rice algorithm.

Additional tasks

Below are **suggested** additional tasks to obtain good or excellent grades for the project. We want to emphasize that students are encouraged to come up with their own ideas for additional investigations and not all of the suggestions below must be included for an excellent grade.

- Ensure that eq. (1) is fulfilled in your model output.
- Use eq. (2) for data augmentation. Does this improve model performance?
- Quantify model uncertainty
- Hyper-parameter tuning
- Visualization of higher dimensional equilibrium data

We want to emphasize that the above points are only suggestions for potential further research questions.

Deliverables

The following materials have to be submitted in order to pass the semester project:

- **Recorded final presentation** (video screencast). The presentation must be **5-7 minutes** (for the entire group) and the file should not exceed **200 mb**. Highlight on the slides which group member(s) are responsible.

- **Written report** to present and discuss the obtained results. You must use the supplied template on Moodle and write no more than **3-4 pages** (for the entire group). Highlight which group member worked on which section.
- **Source code** of your project. Please ensure that the code is executable and optionally add a short explanation of the structure (readme).

Please ensure that all formal conditions (e.g. page limits, highlight responsible author) are satisfied, as we will deduct points for significant violations. Please submit all deliverables via moodle.

Peer review

Part of your grade is to submit a peer review of another project. Peer reviewing will be available on Moodle after submission of the project. Please keep the peer reviewing process in mind when creating your final presentation and report. Your supplied material must be sufficient for other students to understand your work.

Responsible tutor

Please address questions to:

Name	Contact
Benjamin Karg	benjamin.karg@tu-dortmund.de

References

[1] Joachim Gross and Gabriele Sadowski. “Perturbed-chain SAFT: An equation of state based on a perturbation theory for chain molecules”. In: *Industrial & engineering chemistry research* 40.4 (2001), pp. 1244–1260.