LABORATORY FOR CHEMICAL TECHNOLOGY
Technologiepark 125 9052 Gent, Belgium

Coach	Supervisor(s)	Funding
Ir. César Pernalete	Dr. Ir. Florence Vermeire Prof. Dr. Ir. Joris Thybaut Prof. Dr. Ir. Kevin van Geem.	-

A study of the impact of uncertainty prediction methods in active machine learning assisted Design of Experiments for Hydrocracking

Machine Learning, Uncertainty Predictions, Design of Experiments, Hydrocracking

Aim

The aim of this thesis is to evaluate the impact of different methods for the estimation of uncertainty in machine learning models predictions in the context of the design of experiments for the hydrocracking at bench scale of plastic derived hydrocarbon mixtures.

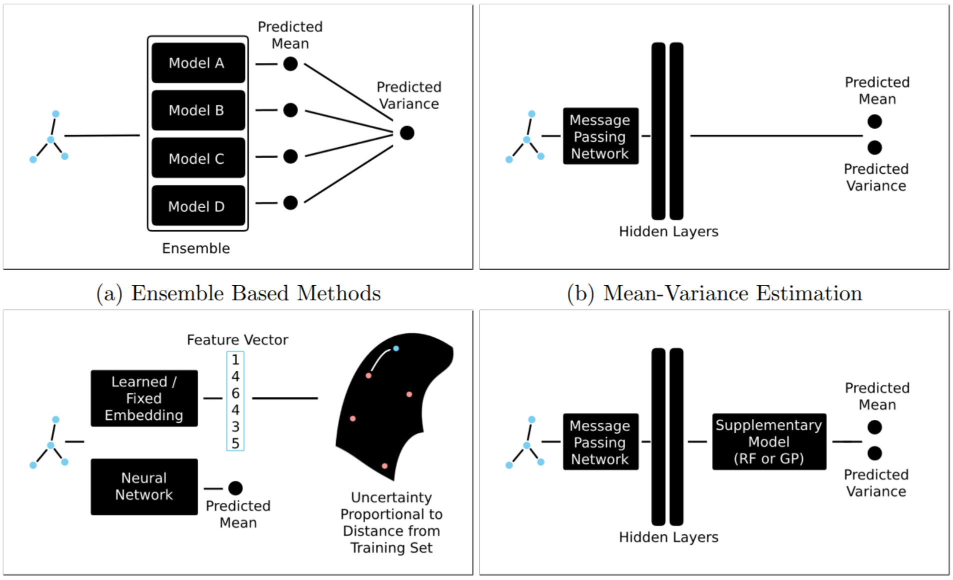
Justification

The hydrocarbon industry is playing a central role in establishing a circular economy. Including pyrolysis oil (PO) from post-consumer plastics into conventional conversion processes, such as hydrocracking (HCK), represents a desirable chemical recycling route. This option offers significant flexibility because it enables the production of a various chemicals including feedstocks for plastics.

This work aims to acquire a deeper understanding of the HDC performance when processing PO from post-consumer plastics. For this purpose, an approach that combines fundamental modelling and a model-driven experimental campaign will be applied. The experiments comprise the processing of mixtures of a fossil derived hydrocarbon mixture with PO under typical hydroconversion conditions and catalysts [1].

The experimental conditions will be selected using an active learning based experimental design (DoE) strategy on which an initial machine learning model trained on existing experimental data. This model is later used to determine the points on which the prediction of 'unlabeled data' have the higher uncertainty. On the one hand, one of the issues to be addressed in this research is the variety of existing options for uncertainty estimation. For that reason different methods for uncertainty estimation will be assessed (ensembling, conformal prediction, e.g.) (Figure 1) [2]. On the other hand, since large amounts of data are required to train the initial machine learning models, in every case the used machine learning model will exploit the capabilities of available fundamental models using different strategies depending on the case (transfer learning, meta-models, e.g.) [3].

The results obtained in the experimentation will be used in two directions. On the one hand the experimental results will be used to re-train the Machine Learning models ensuring that it is progressively better adjusted to the real system. On the other hand, the data will be used to further validate the fundamental model.



(a) Ensemble Based Methods
(b) Mean-Variance Estimation
(c) Distance Based Methods
(d) Union Based Methods

Figure 1: Illustration of different uncertainty estimator [2]

Program

1. Literature review comprising:
    a. active learning based design of experiments
    b. uncertainty estimations in machine learning model predictions
2. Design of experiments based on active learning using different approaches for uncertainty estimations
3. Perform the experimental campaign using a Robinson-Mahoney type reactor.
    a. Analyze feed and product samples using 2D chromatographic analysis.
    b. Calculate the mass balance for each of the experiments.
4. Interpret the obtained results and make conclusions about the different used uncertainty estimation methods.

References

[1] Pernalete, C et. al (2021). The role of hydroconversion in a circular economy: processing of fossil and polymer-based mixtures.
[2] Hirschfeld, L; et. al. (2020). Uncertainty Quantification Using Neural Networks for Molecular Property Prediction.
[3] Vermeire, F and Green, W. H (2021), Transfer learning for solvation free energies: From quantum chemistry to experiments.