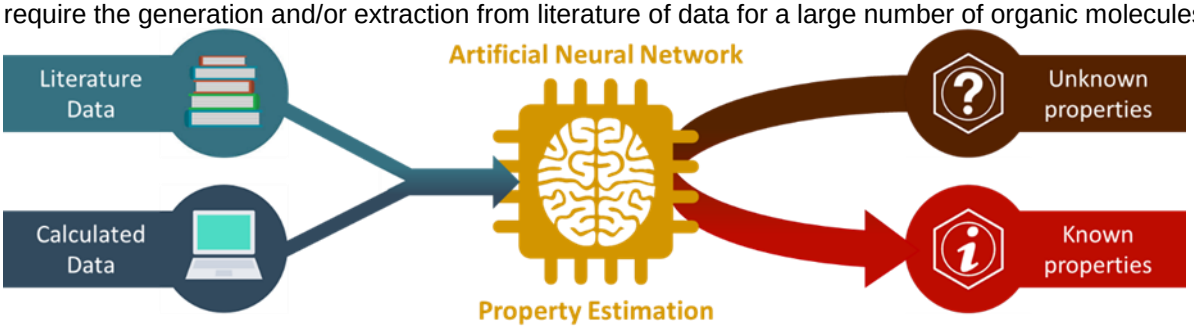
LABORATORY FOR CHEMICAL TECHNOLOGY
Technologiepark 125, 9052 Gent, Belgium

Coach	Supervisor(s)	Funding
dr. ir. Florence Vermeire	prof. dr. ir. Kevin Van Geem, dr. ir. Ruben Van De Vijver	

Big Data Meets Chemistry – Neural Network Based Property Estimation

Aim

The aim of this master thesis is to develop models for the prediction of various thermodynamic and physical properties such as enthalpies and entropies of formation, and solvation energies. These models can then be used in kinetic model generators to predict reaction rate parameters such as activation energies and pre-exponential factors. Depending on the used data, the model could also be used to predict yields and product concentrations in a retrosynthesis tool. Development of these models will require the generation and/or extraction from literature of data for a large number of organic molecules.



Justification

Detailed kinetic models are essential in the development and optimization of various chemical processes. Currently, kinetic model generation tools such as Genesys [1, 2] and RMG [3, 4] rely on group additivity [5] and libraries to estimate reaction rate coefficients for the various encountered reactions. While this has proven to be an efficient and fairly accurate estimation method, there are still several drawbacks. One is the necessity of a reference reaction for each set of group additive values. This increases the complexity of the group additive database and requires some basic user experience when generating the input. Another issue is that of data consistency. When using group additivity it is important that all the used data has similar sources or has been calculated at the same level of theory. Hence the obtained values depend on the data source. These problems can be circumvented by using a less strict estimation approach. Neural networks can handle data from various sources. Moreover, assuming that no single source is 100 % accurate, having several data points for one reaction can even improve the overall prediction accuracy of the network. With vast amounts of available reaction data – both experimental and *ab initio* – it is possible to train a well-designed neural network to learn certain phenomena and their contributions towards the kinetics of a reaction. Examples are ring strain corrections and non-nearest neighbour effects. In the group additivity approach, additional correction factors have to be determined to account for these effects. These often require additional efforts and are not always easily generalized. Using a neural network for the estimation of kinetic parameters fits in the context of developing a self-learning kinetics estimator.

Program

- Literature survey on existing Artificial Intelligence-based methods for property prediction
- Generation of clean datasets of thermo-physical properties of a wide range of organic molecules.
- Potentially supplement (literature) data with ab-initio (or other) calculated values.
- Draw up the architecture for and training of an artificial neural network for the prediction of thermo-physical properties.

1. Van de Vijver, R., et al., *Chemical Engineering Journal* **2015,** 278, 385-393.
2. Vandewiele, N. M., et al., *Chemical Engineering Journal* **2012,** 207–208, 526-538.
3. Gao, C. W., et al., *Computer Physics Communications* **2016,** 203, 212-225.
4. Gao, C. W., et al., *Combustion and Flame* **2015,** 162, (8), 3115-3129.
5. Benson, S. W., In *Berichte der Bunsengesellschaft für physikalische Chemie*, 2nd Ed. ed.; Wiley-VCH Verlag GmbH & Co. KGaA: 1977; Vol. 81, p 320.