Project 1: Machine Learning to Estimate Standard Enthalpy of Formation of Hydrocarbons

Introduction

Chemical kinetics and molecular thermochemistry studies are of vital importance for the development of kinetic models, which are useful to gain knowledge of the oxidation properties of fuels. Accurate thermochemical properties are critical for combustion modeling in the design of chemical process. Some computational methods have been proposed to estimate standard enthalpy of formation, including quantum chemistry calculations and group additive estimation. Recently, with more and more open data available to research community, machine learning methods have been proposed as an alternative method to estimate standard enthalpy of formation.

In this project, you will use the data published in the paper "Machine Learning To Predict Standard Enthalpy of Formation of Hydrocarbons" [1] to estimate Enthalpy of Formation Reaction of hydrocarbons. Machine learning models are powerful to determine the quantitative structure-property relation (QSPR). You can build several basic ML models and compare the accuracy of estimation with the paper.

Procedure:

1. Extract data from the paper [1] to create a data table;

2. Generate SMILES for the molecules;

3. Call RDKit package to generate molecule descriptors.

4. Implement machine learning models to fit and predict standard enthalpy of formation with the selected features.

Reference:

1. "Machine Learning To Predict Standard Enthalpy of Formation of Hydrocarbons", Goergi St. Cholakov, Kiran K. Yalamanchi, Vincent C. O. van Oudenhoven, Francesco Tutino, M. Monge-Palacios, Abdulelah Alshehri, Xin Gao, and S. Mani Sarathy, *J. Phys. Chem. A 2019, 123, 38, 8305–8313.*