DTU Physics
Department of Physics

DTU

Predicting Heat of Formation for 2D Materials Using Machine Learning

M. Petersen (s214452), C. Toft (s214450), E. Voorrips (s243085)

1 Introduction

The aim of this project is to predict the heat of formation of 2D materials using machine learning. More specifically, a Gaussian Process model was used with the Smooth Overlap of Atomic Positions (SOAP) as a fingerprint. Once the model was developed, it was applied in a competition to evaluate how accurately it predicted the heat of formation of 2D materials compared to other groups in the 10316 course. We were given a training set of 8000 compounds with DFT-calculated heats of formation, as well as a test set of 4000 compounds used for the competition

2 Gaussian process

A Gaussian Process is a probabilistic model used in machine learning to learn unknown functions from data. It defines a distribution over possible functions rather than producing a single prediction, allowing both a mean estimate and uncertainty to be quantified. A prior distribution over functions is defined using a mean function and a kernel, which describe the relationship between input points before any data is observed. In this work, a squared exponential kernel is used and shown in equation 1. The prediction is shown in equation 2 with the matrix element that includes the noise 3.

$$k(x, x') = k_0 \cdot exp(-(x-x')^2/2l^2) \quad (1)$$

$$y_0(x) = \mathbf{k}(\mathbf{x})^T \mathbf{C}^{-1} \mathbf{t} \quad (2)$$

$$\mathbf{C} = \mathbf{K} + \sigma \mathbf{I}_N \quad (3)$$

3 Fingerprint: Smooth Overlap of Atomic Positions

A fingerprint is a way to describe features of objects in a mathematical way and makes it possible to compare objects to each other. In material science the fingerprint is most often a representation of the position and chemical composition of the compounds present. Smooth overlap of atomic positions (SOAP) is a fingerprint which represents the local environment of an atom. The dimensions of the SOAP fingerprint depend on the number of different atoms in the data, in this case 62. Leading to fingerprints of 862 792 dimensions. Fingerprint memory usage was minimized from 55 GB to 250 MB by using sparse matrices.

4 Principal Component Analysis PCA

Reduction of the dimensions is necessary when working on large and complex fingerprints. PCA finds orthogonal directions in order of how well they explain the variance in the data. With a reduction from 862 792 to 200 dimensions 91.98% of the variance was still explained.

5 Neural Network

A neural network was used to examine improvements compared to the optimized, but lacking Gaussian Process. We used the SchNet model, a convolutional neural network model developed specifically for quantum chemical interactions [1]. It uses continuous convolutional layers, which allows it to handle unevenly spaced data such as the distribution of atomic positions.

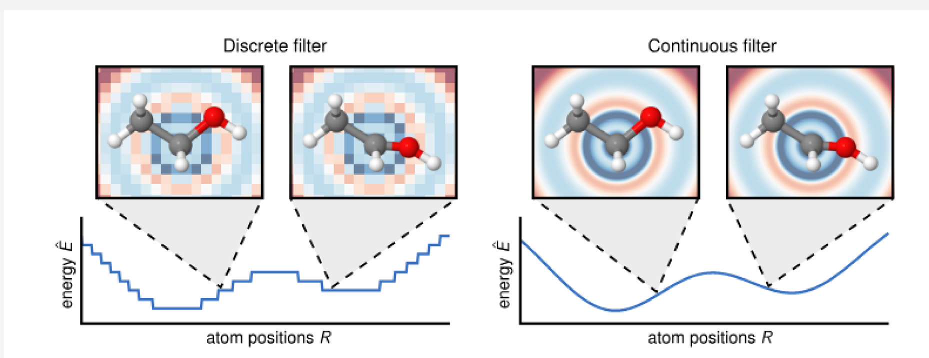

Figure 1: Example of continuous filters compared to discrete filters [1].

This model pushed the results from RMSE≈0.30 for the GP to RMSE≈0.18. Giving the model information about Periodic Boundary Conditions (useful for crystals) and global features such as density and volume quickened learning and pushed the RMSE scores down. The largest increase in performance came due to averaging of predictions for an ensemble of trained models.

6 Results

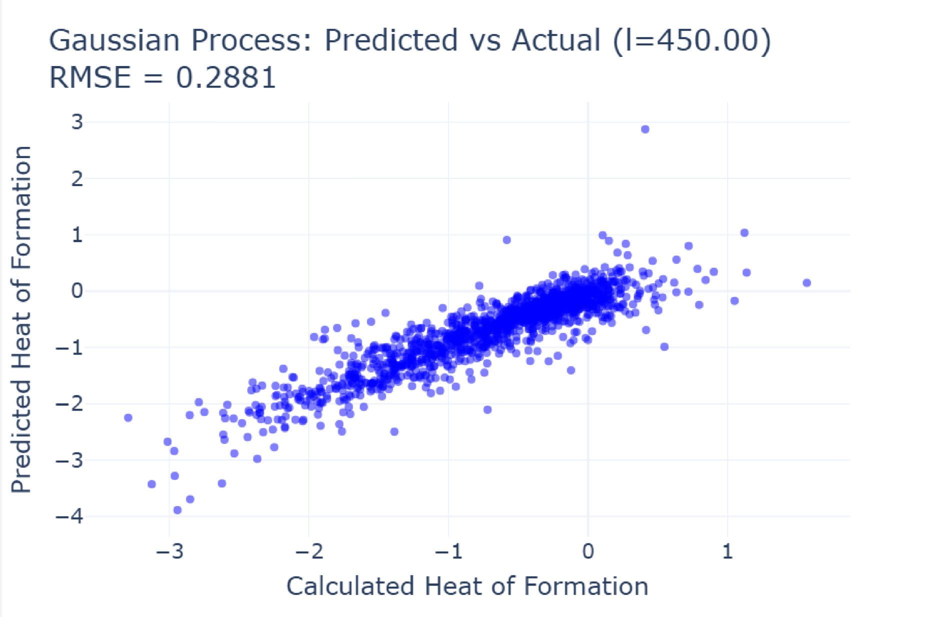

Figure 2: RMSE of the Gaussian proces of the predicted values of the heat of formation compared to calculated values done with DFT.

Model	Description	Lowest RMSE
GP	Coulomb Matrix	0.607
GP	SOAP	0.293
SchNet	standard	0.188
SchNet	incl. PCB	0.168
SchNet	PCB, global features	0.150
SchNet	ensemble (N=10)	0.142

Table 1: Performances of SchNet Neural Network and GP trained on the dataset, tested

7 Discussion

Further optimization could be done for the Gaussian Process, for example by using more accurate hyperparameters. In general, a more accurate model could be achieved by using pre-trained massive models and tailoring it to our dataset. An accurate prediction will help in determining stable compounds that could be suitable for use as 2D materials.

References

[1] Kristof T. Schütt, Pieter-Jan Kindermans, Huziel E. Sauceda, Stefan Chmiela, Alexandre Tkatchenko, and Klaus-Robert Müller. Schnet: A continuous-filter convolutional neural network for modeling quantum interactions, 2017.