Predicting Nonlinear Behaviour in Multidimensional Space with Unsupervised and Supervised Machine Learning

Student: Allan Ng

Project Sponsor: Dr. Mauricio Ponga

Project Period: September 2018 – April 2019

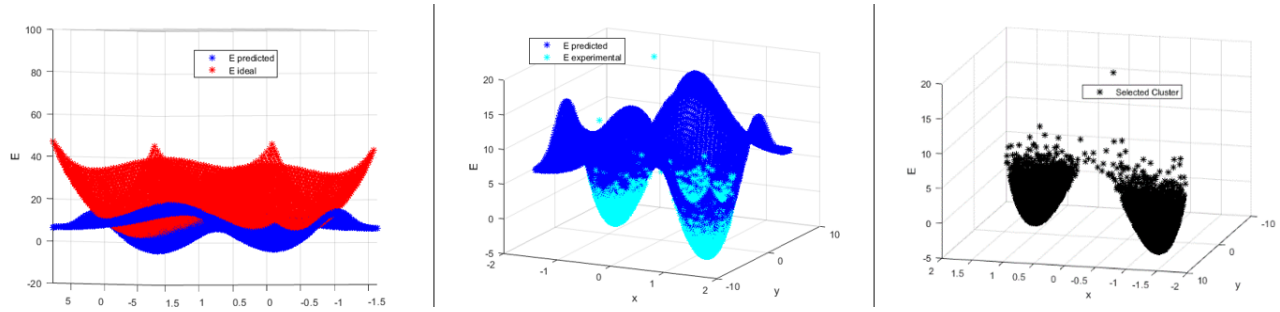

Figure 1: An implementation of unsupervised learning to cluster a large dataset into pertinent piecewise components (right), supervised learning to generate a predicted model which is then compared against the original experimental data and an ideal behaviour test case (middle and left, respectively).

Material behaviour, specifically the energy landscape of graphene sheets, can behave nonlinearly over the span of a sample. Graphene sheets have geometric nonlinearities – peaks and wells scattered across the sheet which make it difficult to take point measurements of the energy at discretized nodes. This introduces a dataset with high and low sample density, and an unevenly defined strain-energy landscape. Because the data in the low sample density regions is not enough to describe the behaviour of the material in those zones, regression (a form of supervised machine learning) is proposed to predict the missing energy response variables. However, for scalability into any material sample with unknown behaviour, the machine learning application is developed to cycle through all nineteen of MATLAB's built in regression models and optimize for the best performing model. To improve the efficiency of the program, K-Means Learning clustering (a form of unsupervised learning) is used to identify independent sub-function representations of the overall nonlinear system to run regression only in the region where the desired response is sought after. This process reduces the computational effort of the program on the premise that nonlinear material behaviour can be modeled as a continuous summation of multiple piecewise sub-functions. This study develops a proof of concept method that shows potential for optimization of the method to produce accurate predicted models. Currently, the un-clustered system still reports the most accurate model, but the clustered model shows a quicker rate of improvement with increased size of datasets compared to the un-clustered system, as well as a slower rate of increase in run time compared to the un-clustered system as input dataset size increases.

☐ Copyright held by UBC. UBC Authors are: Allan Ng, Mauricio Ponga

OR

☐ The copyright holders of this file grants permission for its use through a Creative Commons license, and the uploading and use of this file to the UBC Wiki is compliant with the license terms (select specific license):

- ☐ Creative Commons Attribution 2.5 Canada (legal code)
- ☐ Creative Commons Attribution ShareAlike 3.0 (legal code)
- ☐ Creative Commons Attribution ShareAlike 2.5 (legal code)
- ☐ Creative Commons Attribution 3.0 (legal code)
- ☐ Creative Commons Attribution 2.5 (legal code)
- ☐ Creative Commons CC0 Waiver (release all rights, like public domain: legal code