

Uncertainty Benchmark (placeholder)

Kevin Tran,^{†,¶} Willie Neiswanger,^{‡,¶} Junwoong Yoon,[†] Eric Xing,[‡] and Zachary
W. Ulissi^{*,†}

[†]*Department of Chemical Engineering, Carnegie Mellon University, Pittsburgh, PA 15217*

[‡]*Department of Machine Learning, Carnegie Mellon University, Pittsburgh, PA 15217*

[¶]*These authors contributed equally to this work*

E-mail: zulissi@andrew.cmu.edu

Abstract

Abstract here.

Introduction

1. ML/DS + catalysis
2. Why uncertainty?
 - (a) want confidence on DFT predictions themselves
 - (b) active routines
3. Very quick overview of the paper (Figure 1)



Figure 1: Placeholder for overview of the paper

Methods

Data handling

All regressions in this paper were performed on a dataset of density functional theory (DFT) calculated adsorption energies created with the Generalized Adsorption Simulator for Python (GASpy).^{1,2} These data included energies from 21,269 different H adsorption sites; 1,594 N sites; 18,437 CO sites; 2,515 O sites; and 3,464 OH sites; totaling in 47,279 data points. GASpy performed all DFT calculations using the Vienna Ab-initio Simulation Package (VASP)³⁻⁶ version 5.4 implemented in the Atomic Simulation Environment (ASE).⁷ The revised Perdew-Burke-Ernzerhof (rPBE) functionals⁸ were used along with VASP’s pseudopotentials, and no spin magnetism or dispersion corrections were used. Bulk relaxations were performed with a $10 \times 10 \times 10$ k-point grid and a 500 eV cutoff, and only isotropic relaxation were allowed during this bulk relaxation. Slab relaxations were performed with k-point grids of $4 \times 4 \times 1$ and a 350 eV cutoff. Slabs were replicated in the X/Y directions

so that each cell was at least 4.5 Å wide, which reduces adsorbate self-interaction. Slabs were also replicated in the Z direction until they were at least 7 Å thick, and at least 20 Å of vacuum was included in between slabs. The bottom layers of each slab were fixed and defined as those atoms more than 3 Å from the top of the surface in the scaled Z direction.

To split the data into train/validate/test sets, we first enumerated all adsorption energies on monometallic slabs and added them to the training set manually. We did this because some of the regression methods in this paper use a featurization that contains our monometallic adsorption energy data,¹ and so having the monometallic adsorption energies pre-allocated in the training set prevented any information leakage between the training set and validation/test sets. After this allocation, we performed a 64/14/20 train/validate/test split that was stratified⁹ by adsorbate. We used the train/validate partitions to tune various hyperparameters manually. To create the results shown in this paper, we combined the training and validation partitions into a single training set and reported models’ performances on the held-out training set.

1. Modeling

- (a) CGCNN
- (b) CGCNN Ensemble
- (c) GP
- (d) GP with CGCNN
- (e) With other kernels too
- (f) Penultimate-Fed GP
- (g) Bayesian CGCNN with prior on weights at some layer
- (h) Supervised error prediction (delta CGCNN)
- (i) Dropout CGCNN

2. Assessment

- (a) accuracy
- (b) calibration
- (c) sharpness

Results

1. Table/figure of accuracies: MSE, MAE, R2, [willie get list]
2. Plots:
 - (a) Parity plots
 - (b) Calibration/sharpness plots
 - (c) Sharpness values per method
3. Blocking results?
4. Cost of computing each method (if its there)
5. Human overhead and difficulty

Conclusions

Observations about relative accuracies, calibrations, sharpnesses, overhead

Code availability

Visit https://github.com/ulissigroup/uncertainty_benchmarking for the code used to create the results discussed in this paper. The code dependencies are listed inside the repository.

Author information

Corresponding author email: zulissi@andrew.cmu.edu. The authors declare no competing financial interest.

Acknowledgements

This research used resources of the National Energy Research Scientific Computing Center, a DOE Office of Science User Facility supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.

References

- (1) Tran, K.; Ulissi, Z. W. *Nature Catalysis* **2018**, *1*, 696–703.
- (2) Tran, K.; Aini, P.; Back, S.; Ulissi, Z. W. *Journal of Chemical Information and Modeling* **2018**,
- (3) Kresse, G.; Hafner, J. *Physical Review B* **1993**, *47*, 558–561.
- (4) Kresse, G.; Hafner, J. *Physical Review B* **1994**, *49*, 14251–14269.
- (5) Kresse, G.; Furthmüller, J. *Computational Materials Science* **1996**, *6*, 15–50.
- (6) Kresse, G.; Furthmüller, J. *Physical Review B* **1996**, *54*, 11169–11186.
- (7) Hjorth Larsen, A. et al. *Journal of Physics: Condensed Matter* **2017**, *29*, 273002.
- (8) Hammer, B.; Hansen, L. B.; Nørskov, J. *Physical Review B* **1999**, *59*, 7413–7421.
- (9) Thompson, S. K. In *Sampling*, 3rd ed.; Shewhart, W. A., Wilks, S. S., Eds.; John Wiley and Sons Inc., 2012; Chapter 11, pp 139–156.