Uncertainty Benchmark (placeholder)

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

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

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

¶*These authors contributed equally to this work*

E-mail: zulissi@andrew.cmu.edu

Abstract

Abstract here.

Introduction

The fields of catalysis and materials science are burgeoning with research articles that use machine learning and data science to screen, design, and understand materials.[1–4] This research has spurned the creation of machine learning models to predict various material properties. Unfortunately, the design spaces for these models are sometimes too large and intractable to sample completely. These undersampling issues can limit the training data and therefore the predictive power of the models. It would be helpful to have an uncertainty quantification (UQ) for a model so that we know when to trust the predictions and when not to. More specifically: UQ would enable various online, active frameworks for materials discovery and design (e.g., active learning,[5] online active learning,[6] Bayesian optimization,[7] active search,[8] or goal oriented design of experiments[9]).

Such active frameworks have already been used successfully in the field of catalysis and materials informatics. For example: Peterson[10] has used a neural network to perform online active learning of nudged elastic band (NEB) calculations, reducing the number of force calls by an order of magnitude. Torres et al.[11] have used also used online active learning to accelerate NEB calculations, but they used a Gaussian Process (GP) model instead of a neural network. Jinnouchi et al.[12] have used online active learning to accelerate molecular dynamics simulations. Each of these active methods are underpinned by models with UQ, which has garnered increasing attention itself.[13,14]

To our knowledge though, we have not seen many comparisons of different methods for UQ within the field of catalysis and materials informatics. Here we attempt to resolve this issue by benchmarking six different methods for UQ (Figure 1). We acknowledge that there will not be one optimal method across all use cases, but we still find value in sharing these results so that others can build intuition from our results. Perhaps more importantly, we have also established a protocol for comparing the performance of different modeling and UQ methods.

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).[15,16] 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)[17–20] version 5.4 implemented in the Atomic Simulation Environment (ASE).[21] The revised Perdew-Burke-Ernzerhof (rPBE) functionals[22] were used along with VASP's pseudopotentials, and no spin magnetism or dispersion corrections were used. Bulk

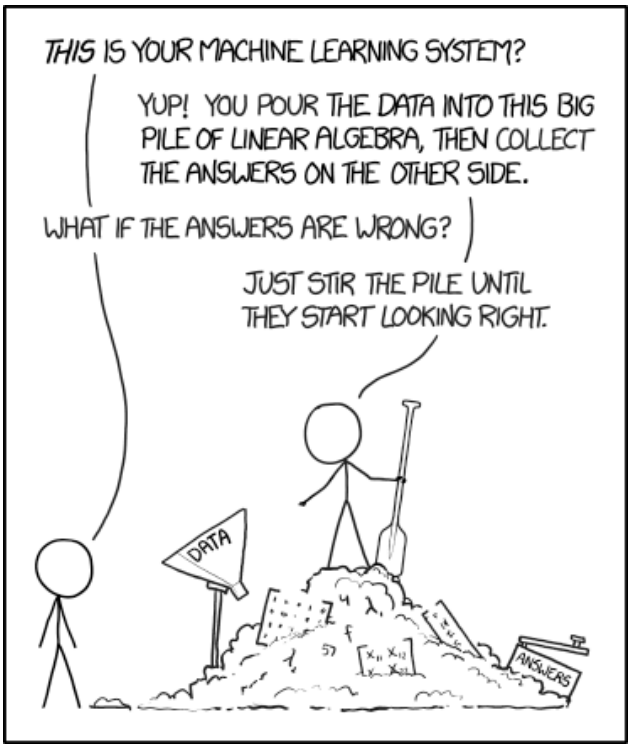

Figure 1: Placeholder for overview of the paper

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,[15] 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 stratified[23] by adsorbate. We then used the validation set's results to tune

various hyperparameters manually. After tuning, we calculated the training set results and present them in this paper exclusively.

Regression methods

1. CGCNN
2. CGCNN Ensemble
3. GP
4. GP with CGCNN
5. With other kernels too
6. Penultimate-Fed GP
7. Bayesian CGCNN with prior on weights at some layer
8. Supervised error prediction (delta CGCNN)
9. Dropout CGCNN

Performance metrics

1. accuracy
2. calibration
3. 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) Medford, A. J.; Kunz, M. R.; Ewing, S. M.; Borders, T.; Fushimi, R. *ACS Catalysis* **2018**, *8*, 7403–7429.

(2) Gu, G. H.; Noh, J.; Kim, I.; Jung, Y. *Journal of Materials Chemistry A* **2019**, *7*, 17096–17117.

(3) Schleder, G. R.; Padilha, A. C. M.; Acosta, C. M.; Costa, M.; Fazzio, A. *Journal of Physics: Materials* **2019**, *2*, 1–46.

(4) Alberi, K. et al. *J. Phys. D: Appl. Phys* **2019**, *52*, 1–48.

(5) Settles, B. In *Synthesis Lectures on Artificial Intelligence and Machine Learning*; Brachman, R. J., Cohen, W. W., Dietterich, T. G., Eds.; Morgan & Claypool, 2012; p 100.

(6) Chu, W.; Zinkevich, M.; Li, L.; Thomas, A.; Tseng, B. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* **2011**, 195–203.

(7) Frazier, P. I. **2018**,

(8) Garnett, R.; Krishnamurthy, Y.; Xiong, X.; Schneider, J.; Mann, R. *Proceedings of the 29th International Conference on Machine Learning, ICML 2012* **2012**, *2*, 1239–1246.

(9) Kandasamy, K.; Neiswanger, W.; Zhang, R.; Krishnamurthy, A.; Schneider, J.; Poczos, B. **2018**,

(10) Peterson, A. A. *Journal of Chemical Physics* **2016**, *145*.

(11) Torres, J. A. G.; Jennings, P. C.; Hansen, M. H.; Boes, J. R.; Bligaard, T. **2018**, *156001*, 1–6.

(12) Jinnouchi, R.; Lahnsteiner, J.; Karsai, F.; Kresse, G.; Bokdam, M. *Physical Review Letters* **2019**, *122*, 225701.

(13) Peterson, A. A.; Christensen, R.; Khorshidi, A. *Phys. Chem. Chem. Phys. Phys. Chem. Chem. Phys* **2017**, *19*, 10978–10985.

(14) Musil, F.; Willatt, M. J.; Langovoy, M. A.; Ceriotti, M. *Journal of Chemical Theory and Computation* **2019**, *15*, 906–915.

(15) Tran, K.; Ulissi, Z. W. *Nature Catalysis* **2018**, *1*, 696–703.

(16) Tran, K.; Aini, P.; Back, S.; Ulissi, Z. W. *Journal of Chemical Information and Modeling* **2018**,

(17) Kresse, G.; Hafner, J. *Physical Review B* **1993**, *47*, 558–561.

(18) Kresse, G.; Hafner, J. *Physical Review B* **1994**, *49*, 14251–14269.

(19) Kresse, G.; Furthmüller, J. *Computational Materials Science* **1996**, *6*, 15–50.

(20) Kresse, G.; Furthmüller, J. *Physical Review B* **1996**, *54*, 11169–11186.

(21) Hjorth Larsen, A. et al. *Journal of Physics: Condensed Matter* **2017**, *29*, 273002.

(22) Hammer, B.; Hansen, L. B.; Nørskov, J. *Physical Review B* **1999**, *59*, 7413–7421.

(23) 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.