Towards Efficient Uncertainty Estimation in Deep Learning for Robust Energy Prediction in Crystal Materials

Sirui Bi[1], Victor Fung[2], Jiaxin Zhang[3] and Guannan Zhang[3]

[1]Computational Sciences and Engineering Division, Oak Ridge National Laboratory

[2]Center for Nanophase Materials Sciences , Oak Ridge National Laboratory, [3]Computer Science and Mathematics Division, Oak Ridge National Laboratory

Introduction

In the search for advanced materials for various functional applications ranging from energy storage to catalysis, deep learning (DL) has quickly gained traction as a powerful and flexible approach. In the search for advanced materials for various functional applications ranging from energy storage to catalysis, deep learning has quickly gained traction as a powerful and flexible approach.

However, the design space is very large such that we are unable to sample completely for these DL models. The resulting under-sampling challenge can limit the training data and therefore the predictive capability of the models.

Novel contribution

In this work, we aim to conduct an efficient uncertainty estimation in deep learning for conducting robust prediction of material properties using DFT simulation datasets.

- We combine a residual deep architecture as a feature extractor with an approximate Gaussian process (GP) model to efficiently estimate uncertainty using a single forward pass.
- We propose to introduce inducing point GP with fuzzy c-means clustering that is used to represent the full datasets such that we can reduce the computational complexity.
- We show the robust performance of the approach on total energy prediction in a real-world lattice crystal structure SrTiO3 perovskite oxide from material chemistry.

Deep Feature Extractor with Residual Network

Deep kernel learning (DKL) is a well-established approach for estimating uncertainty in deep neural networks with a single forward pass. The overall idea of DKL is to first extract the feature by leveraging deep neural networks and then use the feature extractor as an input to a Gaussian process output which offers probabilistic measurement.

$$\mathscr{K}_{\tau,\theta}(\boldsymbol{x}_i,\boldsymbol{x}_j) \leftarrow \mathscr{K}^*_{\tau}(\Omega_\theta(\boldsymbol{x}_i),\Omega_\theta(\boldsymbol{x}_j))$$

Ω_θ is a deep neural network parameterized by θ

$\mathscr{K}^*_\tau$ is the base kernel

τ Is the hyperparameters of the base kernel

DKL encounters several difficulties, particularly in scaling to large datasets because the exact inference for GP is hampered by the inversion of kernel matrix, where the time complexity scales cubically with the number of data.

Inducing Point GP with Soft Clustering

GP.Titsias (2009) proposed a variational formulation for sparse approximations which jointly infers the kernel parameters τ and inducing inputs by maximizing a lower bound of the log marginal likelihood that is also known as the evidence lower bound, which is defined as

$$L(X) = \log P(\mathbf{X}) - D_{\mathrm{KL}}(Q \parallel P) = -\sum Q(\mathbf{Y}) \log Q(\mathbf{Y}) + \sum Q(\mathbf{Y}) \log P(\mathbf{Y},\mathbf{X})$$

Our approach is built upon this idea with a much smaller number of inducing points $m_s \ll N$ to overcome the scalability issue in GP. Inducing point GP reduces the time complexity of the matrix inversion from $\mathscr{O}(N^3)$ to $\mathscr{O}(m_s^2 N)$.

we propose to use a soft clustering, that is often named as fuzzy c-means (FCM), which aims to minimize the objective function

$$\mathscr{F} = \arg\min_{C} \sum_{i=1}^{N} \sum_{j=1}^{m_s} w_{ij}^{m_f} ||\mathbf{x}_i - \boldsymbol{c}_j||^2, \quad w_{ij} \in [0,1]$$

$C = \{\boldsymbol{c}_1, ..., \boldsymbol{c}_{m_s}\}$ is the cluster centers w_{ij} is the partition matrix

Spectral Normalization as a Regularization

Another limitation in DKL called feature collapse is the uncertainty estimation is sensitive to changes in the input.

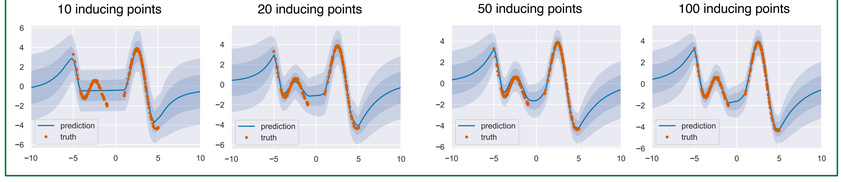
we use spectral normalization as a regularization combined with a residual neural architecture for deep feature extraction.

Algorithm 1: Efficient uncertainty estimation in deep neural networks

1: ***Require***: training data $\{(\mathbf{x}_i, \mathbf{y}_i)_{i=1}^N\}$, wide residual neural networks Ω_θ with parameters θ, the number of inducing point m_s, approximate GP with parameters ξ including inducing point locations ℓ_s, fuzzy hyperparameter, m_f, learning rate λ.
2: ***Initialize inducing points with fuzzy c-means clustering***
3: Draw a random subset of m_r point from the training data $X^{\text{ini}} \subset X$
4: Compute the fuzzy clustering (soft k-means) on $\Omega_\theta(X^{\text{ini}})$ with $k = m_r$, use the centroids as initial inducing point locations ℓ_s in approximate GP.
5: ***Train residual neural networks and GP jointly***
6: Implement spectral normalization on residual neural network parameters $\hat{\theta} \leftarrow \theta$
7: Evaluate forward model to extract the feature space $\psi \leftarrow \Omega_{\hat{\theta}}(\mathbf{x})$
8: Evaluate approximate GP on feature space with parameter ξ, $p(\hat{\mathbf{y}}|\mathbf{x}) \leftarrow \mathrm{GP}_\xi(\psi)$
9: Define the loss function $\mathcal{L}$ using the negative evidence lower bound, $\mathcal{L} \leftarrow \mathrm{NELB}_\xi(p(\hat{\mathbf{y}}|\mathbf{x}), \mathbf{y})$
10: Minimize the loss function $\mathcal{L}$ with respect to θ ξ via $\theta, \xi \leftarrow \theta, \xi + \lambda \cdot \nabla_{\theta,\xi}\mathcal{L}$

Experiments: 1D Sinusoid Regression Problem

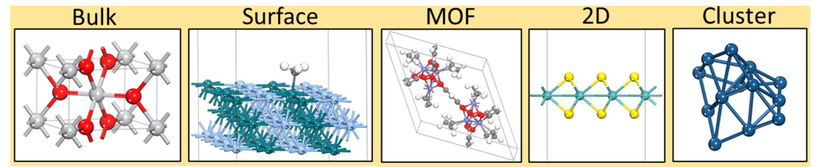
The fewer inducing points (e.g., 10 points) lead to a poor uncertainty estimation and increasing points will improve the prediction including uncertainty as expected



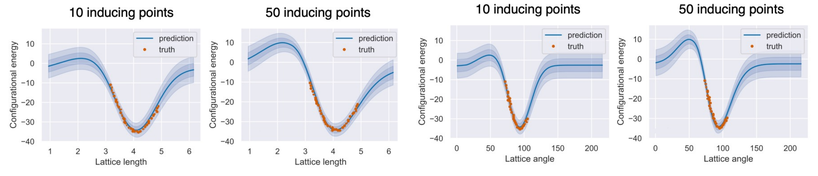
Total Energy Prediction in Crystal Lattice Systems

Materials chemistry simulations are performed to simulate these materials under perturbation and obtain their resulting physical properties such as total energy. Here, we tackle the robust prediction of total energy to strain mapping for the case of the SrTiO3perovskite oxide, which is otherwise intractable to obtain from materials chemistry

This can be an exceptionally difficult problem due to the complex underlying physics, and the high degree of sensitivity of the total energy to the lattice parameters including the length and angle, requiring very accurate predictions for the generated structures to succeed. This proposed method can be extended to study various crystal lattice structures, including bulk, surface, metal-organic frameworks (MOFs), 2D materials, and cluster structures in materials chemistry



The total energy prediction with respect to lattice length and lattice angle parameters given 10 and 50 inducing points



Conclusion and Future Work

We propose an efficient uncertainty estimation in deep learning and apply it for robust prediction of total energy in materials chemistry, specifically crystal lattice systems. Our approach is built upon deep kernel learning (DKL) and addresses the existing challenges by combining spectral normalization and inducing point approximate GP in feature space. **The future work** will compare our approach to other baseline methods, such as Monte Carlo dropout, Bayesian neural networks (BNN), deep ensemble methods and deterministic uncertainty estimation methods in terms of the stability, accuracy and efficiency of predictive uncertainty.

Acknowledgements

This work was supported by the U.S. Department of Energy, Office of Science, Office of Advanced Scientific Computing Research (ASCR), Applied Mathematics program; and by the Artificial Intelligence Initiative at the Oak Ridge National Laboratory (ORNL). This work used resources of the Oak Ridge Leadership Computing Facility, which is supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC05-00OR22725