THE JOURNAL OF PHYSICAL CHEMISTRY A
A JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

pubs.acs.org/JPCA

Article

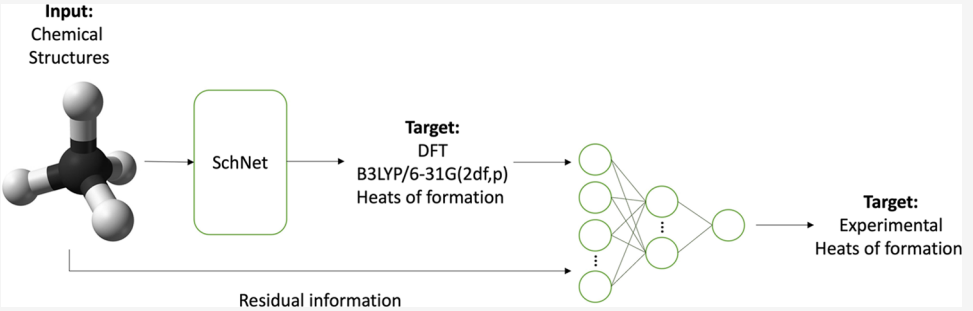
Predicting Experimental Heats of Formation via Deep Learning with Limited Experimental Data

GuanYa Yang,[§] Wai Yuet Chiu,[§] Jiang Wu, Yi Zhou, ShuGuang Chen, WeiJun Zhou, Jiaqi Fan, and GuanHua Chen*

Cite This: *J. Phys. Chem. A* 2022, 126, 6295−6300

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information



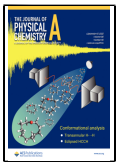
ABSTRACT: When it comes to predicting experimental values of molecular properties with deep learning, the key problem is the lack of sufficient experimental data for training. We propose a method that consists of pretraining a graph neural network that aims to reproduce first-principles quantum mechanical results, followed by fine-tuning of a fully connected neural network against experimental results. The combined pretraining and fine-tuning model is expected to yield molecular properties close to experimental accuracy. This is made possible because first-principles quantum mechanical methods are often qualitatively correct or semiquantitatively accurate; thus, a calibration of the calculation results against high-precision but limited experiment data can improve accuracy greatly. Moreover, the method is highly efficient, as first-principles quantum mechanical calculation is bypassed. To demonstrate this, we apply the combined model to determine the experimental heats of formation of organic molecules made of H, C, O, N, or F atoms (up to 30 atoms), where mere 405 experimental data are used. The overall mean absolute error is 1.8 kcal/mol for these molecules.

INTRODUCTION

Graph neural networks (GNNs)[1] have been employed to predict quantum mechanical properties of molecules.[2,3] They include message passing neural networks[2] and dense tensor neural networks.[3] A recent review[4] summarizes possible applications of GNNs, including learning molecular fingerprints, molecular optimization, and molecular generation. Molecules can be intuitively considered as graphs by chemists, as atoms being nodes and bonds being edges. GNNs are designed to be input with structural information of any molecule and output the molecular properties of interest. Most of these studies[5] predict the molecular properties calculated at the density functional theory (DFT)[6] level. As the solution of the Kohn−Sham equation is bypassed, the computational efficiency can be greatly improved.

However, DFT results are often not accurate enough. Can the accuracy of the GNN-learnt quantum mechanical properties be improved further, for instance, the calculated thermodynamic properties within chemical accuracy compared to experimental results? Direct training of GNNs targeting at experimental values seems to be impractical, since experimental data are expensive to obtain and thus scarce. Limited experimental data are not enough to train a deep network. There have been attempts to employ a reduced amount of data to machine-learning molecular properties. Kim and colleagues[7] proposed a self-supervised network to learn molecular representations from the Simplified Molecular-Input Line-Entry System, followed by a fine-tuning with smaller datasets (ranging from 600 to 90,000 molecules). Moreover, there was also work exploiting transfer learning with an unsupervised pretraining model to extract structural information and transfer learning with more than 20,000 data to predict carbohydrate reactions.[8] For smaller experimental datasets, simpler network structures are usually used instead. In 2018, a group of researchers predicted solar cell efficiencies with simple

Received: April 28, 2022
Revised: August 23, 2022
Published: September 2, 2022



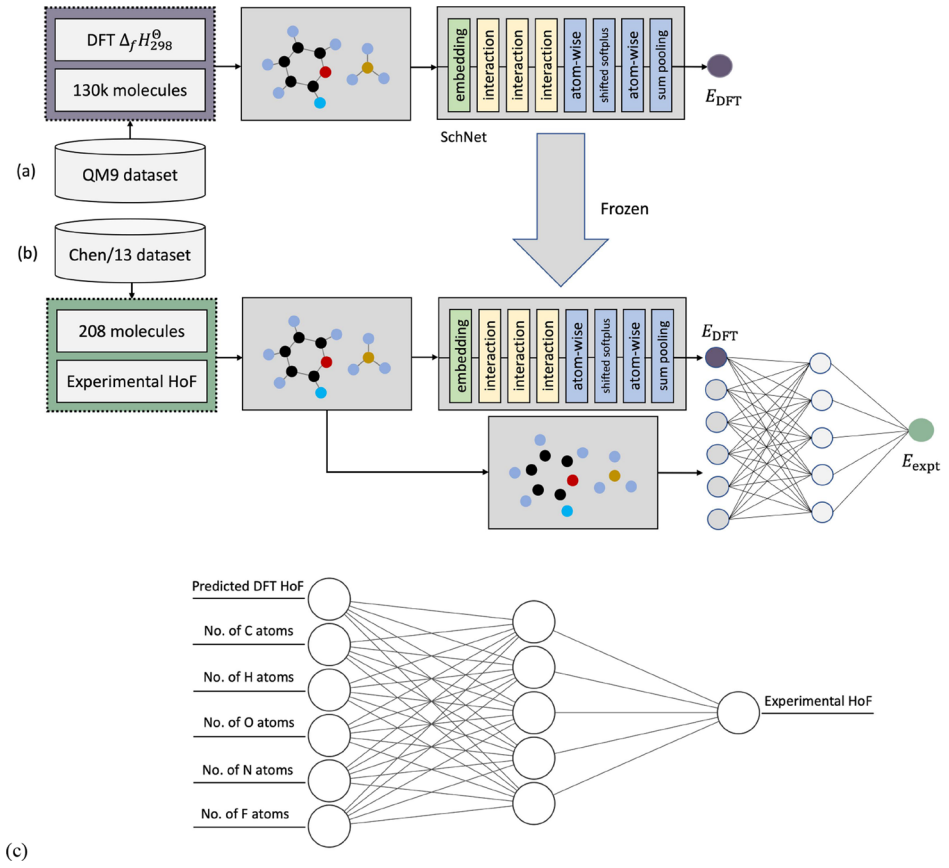


Figure 1. Schematic diagram of the CPTFT approach. It contains a (a) pretraining stage and a (b) fine-tuning stage. (a) In the pretraining step, molecular structural information is fed into SchNet as the input, and DFT HoF, E_{DFT}, is the output. The trained SchNet is frozen and used as part of the fine-tuning. (b) In the fine-tuning step, characteristic molecular properties and E_{DFT} reproduced from the SchNet in (a) are used as the input for the two-layer fully connected NN to predict experimental HoF. The mean absolute error (MAE) for experimental HoFs is 1.8 kcal/mol. (c) Architecture of the two-layer NN. The input layer consists of six input neurons (with one bias neuron, not shown in Figure 1), the hidden layer consists of five neurons, and the output layer consists of one output neuron being the experimental HoF.

machine learning algorithms such as linear regression, the artificial neural network (ANN), and random forest with only 280 experimental data points.[9] Similarly, the ANN was again proven to be capable of predicting semiempirical quantum chemical properties.[10] In one of the most recent studies, a deep-learning ANI-1ccx method was proposed to compute enthalpies of formation of molecules to near chemical accuracy without training directly on experimental values, with time complexity $O(n)$.[11] We are attempting to approach the same level of accuracy and comparable computational cost with another method.

What if we introduce quantum mechanics in between? DFT results are often qualitatively correct or semiquantitatively accurate; consequently, a small and yet accurate experimental dataset would be enough to calibrate DFT results to yield the experimental results. Almost 20 years ago, a calibration method was developed to correct DFT heat of formation (HoF) within chemical accuracy, as compared to the experimental HoFs of organic molecules. A simple two-layer fully connected neural network (NN) and mere 180 accurate experimental values were used for training and testing.[12] The model was further employed or extended in a series of follow-up studies, confirming its reliability.[13] In addition, transfer learning or fine-tuning of NNs has been frequently used in computational chemistry. For example, the Δ-ML model combines fast but approximate quantum mechanical calculation and big data-based machine learning, to reach higher approximation.[14] Also, the discrepancies between DFT-computed properties and experimental measured counterparts can be possibly eliminated by a deep transfer learning approach.[15] As a result, deep learning with fine-tuning can help quantum mechanics in computing molecular properties near experimental accuracy with limited data.

In this study, employing deep learning, we aim to predict experimental HoF values of organic molecules with up to 30 atoms containing C, H, O, N, or F atoms, using molecular structural information only. Rather than predicting directly with a single deep NN, we introduce quantum mechanics in between and employ a combined pretraining and fine-tuning (CPTFT) approach: a deep GNN is constructed and trained to reproduce DFT HoFs at the B3LYP/6-31G(2df,p) level; then, a two-layer fully connected NN is used to further fine-tune the GNN against the experimental HoFs (architecture illustrated in Figure 1). Only limited amounts of experimental data are required for the fine-tuning step (405 HoFs used in our study), while the pretraining step does require a large amount of DFT-calculated HoFs. The network architecture in the fine-tuning step follows the network structure in ref 12, but our overall network prediction requires no expensive DFT calculation, hence more computationally efficient than ref 12.

https://doi.org/10.1021/acs.jpca.2c02957
J. Phys. Chem. A 2022, 126, 6295−6300

METHODS

Construction of the Dataset. The data used for training in this study come from two datasets: QM9[16] dataset and Chen/13 dataset.[17] QM9 is a dataset containing more than 130k stable small organic molecules consisting of C, H, O, N, and F. All molecules within the dataset contain up to nine heavy atoms. The dataset reports calculated quantum chemical properties including energies, enthalpies, and free energies, at the B3LYP/6-31G(2df,p) level. The 539 molecules in the Chen/13 dataset are made up of C, H, O, N, F, S, and Cl, and their experimental HoFs are included. There are 208 molecules, which exist in both datasets; i.e., they have not only DFT HoF values (from QM9) but also experimental HoF values (from Chen/13). They are chosen to be the dataset for fine-tuning and excluded from the training process of GNN.

To pretrain the GNN, 70% molecules from QM9 are randomly selected to be the training set, while 10 and 20% of molecules are used for validating and testing sets. In fine-tuning, the 405 molecules (molecules containing S and Cl, and radicals are excluded from the Chen/13 dataset) of the Chen/13 dataset are randomly split into the training and testing sets with a ratio of 7:3, using the SPXY sampling method and random sampling and KS sampling for comparison.

DFT HoFs are calculated at the B3LYP/6-31G(2df,p) level for molecules in the QM9 dataset (calculation method of HoFs is mentioned in the Supplementary Information). The graph structured data are generated using RDKit and PyTorch Geometric.[18] Atomic attributes are encoded with the atom type, family type (acceptor, donor, or aromatic), hybridization type, and number of neighboring hydrogen atoms. The edge index, which indicates the existence of bonds between atoms, and edge attributes that include the bond type and bond distance are also used as molecular representatives.

Molecular Representation in Graph Neural Networks. Atomic types and positions within a molecule are used when training the SchNet. The position matrix has a shape of $3 \times N$, where N is the number of atoms in the molecule. The atom type matrix is of shape $N \times 1$. The information for methane in matrices form is shown in Figure 2.

$$\text{position} = \begin{pmatrix} -0.0127 & 1.0858 & 0.0080 \\ 0.0022 & -0.0060 & 0.0020 \\ 1.0117 & 1.4638 & 0.0003 \\ -0.5408 & 1.4475 & -0.8766 \\ -0.5238 & 1.4379 & 0.9064 \end{pmatrix}$$

$$\text{atom type} = (6 \quad 1 \quad 1 \quad 1 \quad 1)$$

Figure 2. Matrix representation of methane in the dataset. The position matrix lists the three-dimensional coordinates of the atoms within the molecule. The atom type matrix lists the atomic numbers of all atoms within the molecule.

Computational Details of Graph Neural Networks. Loss function is MSELoss, and the optimizer chosen is Adam.[19] The batch size is 256, and the learning rate is decreased from 0.001 to 0.0005, with the learning rate being halved every 100 epochs.

Computational Details of Two-Layer Neural Networks. The two-layer NN has one hidden layer with five hidden neurons and is fully connected. Bayesian regularization is adopted from a previous study.[20] With Bayesian training, prior and posterior probabilities are considered rather than using fixed values during the updating of weights in backpropagation. This method is designed to identify poor underlying assumptions in NN models and to achieve both good generalizability and Bayesian evidence with a limited size of training data. The objective function can be expressed as follows:

$$J(\hat{y}_i, y_i) = \alpha E_w + \beta E_D$$

where y_i and $\hat{y}_i$ are pairs of experimental heat of formation and the predicted heat of formation, respectively. The objective function J contains two parts: sum of square of weights E_W and sum of square error between the target and predicted value E_D.

Applying Bayer's rule:

$$P(w|D, \alpha, \beta, M) = \frac{P(D|w, \beta, M)P(w|\alpha, M)}{P(D|\alpha, \beta, M)}$$

Given the assumption that weights and noise in the dataset are random variables, NNs can be also explained by Bayes' rule. In the Bayesian NN, α and β are regularization parameters of these two terms. $P(w \mid \alpha, M)$ denotes the prior knowledge of weights before training on any dataset. $P(D \mid w, \beta, M)$ is the probability of data occurring given the weights. $P(D \mid \alpha, \beta, M)$ is used as a normalization constant to guarantee that the total probability is 1.

MALTAB is used to construct a fine-tuning model. The pretrained result from the GNN is used as one of the inputs.

RESULTS

Pretraining: GNN Prediction of DFT Heats of Formation. Molecules are traditionally represented as structural diagrams with bonds and atoms. It is natural to consider molecules as graphs, with atoms being nodes and bonds being edges. A GNN[1] is employed to represent the molecules of interest and trained to produce the DFT HoFs. We use the SchNet[21] to predict DFT HoFs at T = 298 K, $\Delta_f H^{\Theta}_{298}$, for the molecules in the QM9[16] dataset. B3LYP/6-31G(2df,p) is employed to calculate $\Delta_f H^{\Theta}_{298}$ of 134k organic molecules containing up to 30 C, H, O, N, or F atoms. The SchNet is a variant of GNNs. It makes use of continuous-filter convolutional layers to capture interactions among atoms. The output energy is represented as the sum of individual atomic energies.

To investigate the accuracy of the SchNet, the MAEs between predicted and targeted HoFs are calculated and listed in Table 1. The MAEs of SchNet in both training and testing datasets are around 0.2−0.3 kcal/mol.

Table 1. SchNet Predicted Mean Absolute Errors of the HoFs for the Molecules in the Training, Testing, and Validating Datasets

dataset	training set	testing set	validating set
MAE (kcal/mol)	0.22	0.33	0.33

Computation Time Comparison with DFT. In addition to accuracy comparison, the computational efficiencies of both the DFT method and SchNet are also compared. DFT computation is known to have time complexity of $O(n^3)$, which indicates that if the number of electrons within the molecule being computed is doubled, the computational time will be expected to octuple. As for SchNet, the trained NN would not be as sensitive to the size of the molecule as DFT.

https://doi.org/10.1021/acs.jpca.2c02957
J. Phys. Chem. A 2022, 126, 6295−6300

The comparison of computational time between DFT and SchNet is summarized in Figure 3 below.

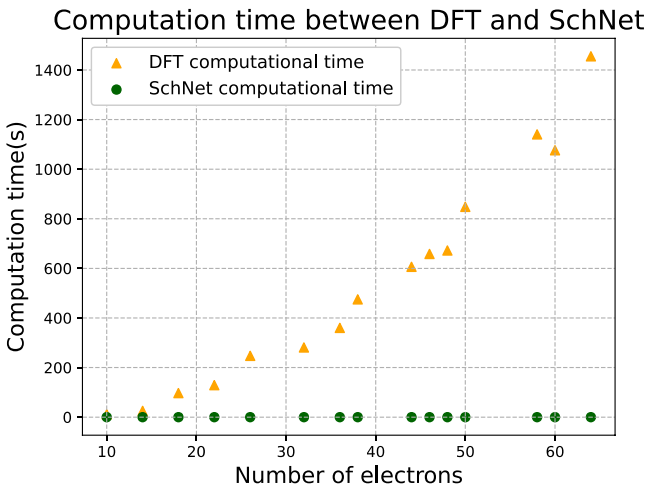


Figure 3. Computational time comparison between DFT and SchNet. Orange triangles indicating computing time for DFT, and dark-green dots indicating SchNet computational time. A significant speed difference is captured in the figure.

Twelve molecules with the electron number ranging from 40 to 64 are selected to be the benchmarking molecules. Their DFT thermodynamic properties are computed using an ORCA[22] single strand, while SchNet is applied on a single GPU. For the molecule with 64 electrons, DFT uses 1455 s for obtaining the result, when SchNet only uses 0.005 s. As a result, SchNet is proved to be much faster than traditional DFT calculation.

Fine-Tuning: Prediction of Experimental Heats of Formation. The classical universal approximation theorem[23] states that the two-layer NN of an arbitrary number of hidden neurons with nonpolynomial activation functions can approximate any continuous functions to arbitrary accuracy. Consequently, it can theoretically calibrate systematic errors between DFT and experimental HoFs. As the number of accurate experimental HoFs is limited, the number of the hidden neurons in the hidden layer must be limited as well as the number of the descriptors. In principle, it is not guaranteed that the two-layer NN with limited hidden neurons/descriptors yields the HoF within the chemical accuracy. Here, we adopt the same procedure and descriptor selection in ref 17 to calibrate further the HoFs resulted from the SchNet against the Chen/13 dataset.

The architecture of the two-layer NN is shown in Figure 1a. The SchNet-output E_{DFT} is taken as the first descriptor to the two-layer NN. Also, 405 experimental HoFs from the Chen/13 dataset are chosen to be the experimental HoF dataset for training and testing. The MAE of the HoFs between the two-layer NN's outputs and the experimental values is around 1.8 kcal/mol (Table 2).

To visualize the performance of the SchNet and two-layer NN, comparisons among various HoFs are shown in Figure 4.

Comparison between Sampling Methods. Training and testing MAEs are compared among three different sampling methods: random sampling, Kennard−Stone (KS) sampling,[24] and sample set partitioning based on joint $x-y$ distances (SPXY) sampling.[25] In the KS and SPXY sampling methods, molecules in the training set are selected furthest away from each other, and the rest of the molecules are in the testing set. The KS sampling method is conducted based on the Euclidian distances of the descriptors between all data points, while SPXY considers distances of not only the descriptors but also the targets. Therefore, they lead to different partitionings of training and testing datasets. In contrast to random sampling, they lead to better testing performances as compared to the training performances. In all three sampling methods, the training and testing ratio are 7:3. The results of all three sampling methods are summarized in Table 3.

Table 2. MAE of SchNet-Produced and CPTFT HoFs for the Training and Testing Molecules against Experimental HoFs

dataset/method	training set (kcal/mol)	testing set (kcal/mol)
SchNet		7.7
CPTFT with SPXY sampling	4.3	1.8

For SPXY, bootstrap analysis[17,20] is performed 100 times. The average value and standard deviation of HoFs are evaluated for each testing molecule selected by SPXY. The error bars, i.e., standard deviation, are plotted for each testing molecule in Figure 4c.

An important issue to be noticed is that the use of both KS and SPXY sampling methods may affect the transferability of the model obtained. However, we performed training and testing on the QM9 dataset and similar molecules, and the effect of the sampling method may be ignored.

DISCUSSION

We have developed the CPTFT approach to predict the experimental HoFs of the organic molecules containing up to 30 C, H, O, N, or F atoms, and only limited experimental data are available for training and testing. This is possible because DFT yields qualitatively correct HoFs of the organic molecules of our interest. DFT/B3LYP is itself efficient enough to calculate sufficient data to pretrain the SchNet to reproduce DFT results, to be specific HoFs in this study. The trained SchNet in the pretraining step is then able to generate more DFT results at a much quicker rate than DFT itself. The fact that DFT calculation is bypassed distinguishes our work from previous work predicting more accurate molecular properties. Then, the pretraining step is followed by fine-tuning the two-layer NN to yield the experimental values. After fine-tuning, systematic errors of DFT results are mostly corrected; however, some random error still exists, which is introduced in the pretraining step. Since our dataset is made of the molecules containing up to 30 atoms, the CPTFT approach cannot be extrapolated to larger molecules. Moreover, it may not be applied to molecules containing atoms other than C, H, O, N, or F.

Besides the HoFs, DFT has been employed to calculate other molecular properties and the results are qualitatively or semiquantitatively correct. This implies that the CPTFT approach can be extended to predict the experimental values of these molecular properties as well. Besides quantum mechanical methods, molecular dynamics,[26] multiscale simulation,[27] coarse grained modeling,[28] finite element methods, and various other methods have been used to calculate the material properties and structures and often yield qualitatively correct results. Our proof-of-principle work suggests that the

https://doi.org/10.1021/acs.jpca.2c02957
J. Phys. Chem. A 2022, 126, 6295−6300

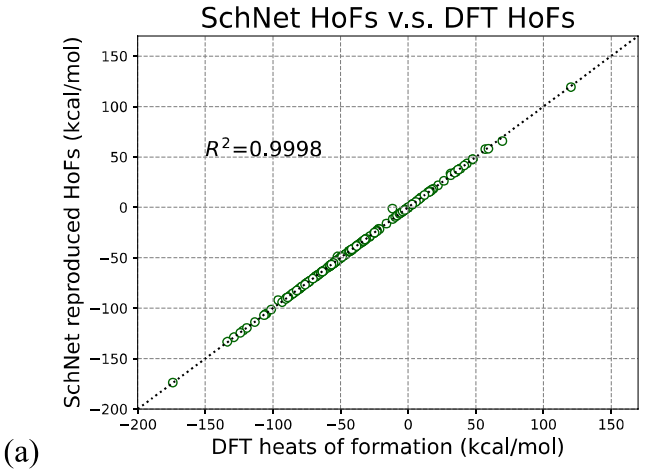


Figure 4. Comparisons among the experimental, DFT, SchNet, and CPTFT HoFs. (a) Comparison between SchNet and DFT HoFs (R^2 = 0.9998). (b) Comparison between SchNet and experimental HoFs (R^2 = 0.9738). (c) CPTFT HoFs versus experimental values (R^2 = 0.9972). Error bars are computed by bootstrapping of testing molecules selected by the SPXY sampling method. Large error bars indicate that the molecule is of less seen structures in the training dataset; therefore, the prediction values vary more.

CPTFT approach can be extended to predict the accurate experimental properties, employing other computational methods besides DFT, and largely reduce the amount of experimental data required.

Table 3. MAE Comparison among Random, KS, and SPXY Samplings

sampling methods	random	KS	SPXY
training MAE (kcal/mol)	3.4	4.3	4.3
testing MAE (kcal/mol)	6.4	2.2	1.8

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jpca.2c02957.

Calculation process of heats of formation for molecules in the QM9 dataset with necessary constants (PDF)

■ AUTHOR INFORMATION

Corresponding Author

GuanHua Chen − *Department of Chemistry, The University of Hong Kong, Hong Kong SAR, China; Hong Kong Quantum AI Lab Limited, Hong Kong SAR, China;* orcid.org/0000-0001-5015-0902; Email: ghc@everest.hku.hk

Authors

GuanYa Yang − *Department of Chemistry, The University of Hong Kong, Hong Kong SAR, China*

Wai Yuet Chiu − *Department of Chemistry, The University of Hong Kong, Hong Kong SAR, China;* orcid.org/0000-0003-0552-8026

Jiang Wu − *Department of Chemistry, The University of Hong Kong, Hong Kong SAR, China; Hong Kong Quantum AI Lab Limited, Hong Kong SAR, China*

Yi Zhou − *Department of Chemistry, The University of Hong Kong, Hong Kong SAR, China;* orcid.org/0000-0002-6183-3060

ShuGuang Chen − *Hong Kong Quantum AI Lab Limited, Hong Kong SAR, China*

WeiJun Zhou − *Department of Chemistry, The University of Hong Kong, Hong Kong SAR, China*

Jiaqi Fan − *Hong Kong Quantum AI Lab Limited, Hong Kong SAR, China*

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jpca.2c02957

Author Contributions

[§]G.Y and W.Y.C. contributed equally.

Notes

The authors declare no competing financial interest.

The code and data that support the findings of this study are available at https://github.com/waiyuetchiu/GNN_finetune_HoF.

■ ACKNOWLEDGMENTS

Financial support from RGC General Research Fund under grant no. 17309620 and Hong Kong Quantum AI Lab Limited is gratefully acknowledged. The authors declare no competing interest. The RGC General Research Fund (17309620) and Hong Kong Quantum AI Lab Limited.

REFERENCES

(1) Scarselli, F.; Gori, M.; Tsoi, A. C.; Hagenbuchner, M.; Monfardini, G. The graph neural network model. *IEEE Trans. Neural Networks* **2009**, *20*, 61−80.

(2) Gilmer, J.; Schoenholz, S. S.; Riley, P. F.; Vinyals, O.; Dahl, G. E. Neural message passing for quantum chemistry. In *International Conference on Machine Learning*; PMLR, 2017; pp 1263−1272.

(3) Schütt, K. T.; Arbabzadah, F.; Chmiela, S.; Müller, K. R.; Tkatchenko, A. Quantum-chemical insights from deep tensor neural networks. *Nat. Commun.* **2017**, *8*, 13890.

(4) Asif, N. A.; Sarker, Y.; Chakrabortty, R. K.; Ryan, M. J.; Ahamed, M. H.; Saha, D. K.; Badal, F. R.; Das, S. K.; Ali, M. F.; Moyeen, S. I. Graph neural network: A comprehensive review on non-euclidean space. *IEEE Access* **2021**, *9*, 60588−60606.

(5) (a) Allotey, J.; Butler, K. T.; Thiyagalingam, J. Entropy-based active learning of graph neural network surrogate models for materials properties. *J. Chem. Phys.* **2021**, *155*, 174116. (b) Hao, Z.; Lu, C.; Huang, Z.; Wang, H.; Hu, Z.; Liu, Q.; Chen, E.; Lee, C. ASGN: An active semi-supervised graph neural network for molecular property prediction. In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*; 2020; pp 731−752. (c) Wu, Z.; Ramsundar, B.; Feinberg, E. N.; Gomes, J.; Geniesse, C.; Pappu, A. S.; Leswing, K.; Pande, V. MoleculeNet: a benchmark for molecular machine learning. *Chem. Sci.* **2018**, *9*, 513−530. (d) Shindo, H.; Matsumoto, Y. *Gated graph recursive neural networks for molecular property prediction*. 2019, arXiv preprint arXiv:1909.00259.

(6) Hohenberg, P.; Kohn, W. Inhomogeneous electron gas. *Phys. Rev.* **1964**, *136*, B864.

(7) Kim, H.; Lee, J.; Ahn, S.; Lee, J. R. A merged molecular representation learning for molecular properties prediction with a web-based service. *Sci. Rep.* **2021**, *11*, 11028.

(8) Pesciullesi, G.; Schwaller, P.; Laino, T.; Reymond, J.-L. Transfer learning enables the molecular transformer to predict regio-and stereoselective reactions on carbohydrates. *Nat. Commun.* **2020**, *11*, 4874.

(9) Sahu, H.; Rao, W.; Troisi, A.; Ma, H. Toward predicting efficiency of organic solar cells via machine learning and improved descriptors. *Adv. Energy Mater.* **2018**, *8*, No. 1801032.

(10) Wan, Z.; Wang, Q. D.; Liang, J. Accurate prediction of standard enthalpy of formation based on semiempirical quantum chemistry methods with artificial neural network and molecular descriptors. *Int. J. Quantum Chem.* **2021**, *121*, No. e26441.

(11) Zheng, P.; Yang, W.; Wu, W.; Isayev, O.; Dral, P. O. Toward Chemical Accuracy in Predicting Enthalpies of Formation with General-Purpose Data-Driven Methods. *J. Phys. Chem. Lett.* **2022**, *13*, 3479−3491.

(12) Hu, L.; Wang, X.; Wong, L.; Chen, G. Combined first-principles calculation and neural-network correction approach for heat of formation. *J. Chem. Phys.* **2003**, *119*, 11501−11507.

(13) (a) Wu, J.; Zhou, Y.; Xu, X. The X 1 family of methods that combines B 3 LYP with neural network corrections for an accurate yet efficient prediction of thermochemistry. *Int. J. Quantum Chem.* **2015**, *115*, 1021−1031. (b) Li, H.; Zhong, Z.; Li, L.; Gao, R.; Cui, J.; Gao, T.; Hu, L. H.; Lu, Y.; Su, Z. M.; Li, H. A cascaded QSAR model for efficient prediction of overall power conversion efficiency of all-organic dye-sensitized solar cells. *J. Comput. Chem.* **2015**, *36*, 1036−1046. (c) Gao, T.; Li, H.; Li, W.; Li, L.; Fang, C.; Li, H.; Hu, L.; Lu, Y.; Su, Z.-M. A machine learning correction for DFT non-covalent interactions based on the S22, S66 and X40 benchmark databases. *J. Cheminf.* **2016**, *8*, 24.

(14) Ramakrishnan, R.; Dral, P. O.; Rupp, M.; von Lilienfeld, O. A. Big data meets quantum chemistry approximations: the Δ-machine learning approach. *J. Chem. Theory Comput.* **2015**, *11*, 2087−2096.

(15) Jha, D.; Choudhary, K.; Tavazza, F.; Liao, W.-K.; Choudhary, A.; Campbell, C.; Agrawal, A. Enhancing materials property prediction by leveraging computational and experimental data using deep transfer learning. *Nat. Commun.* **2019**, *10*, 5316.

(16) (a) Ramakrishnan, R.; Dral, P. O.; Rupp, M.; Von Lilienfeld, O. A. Quantum chemistry structures and properties of 134 kilo molecules. *Sci. Data* **2014**, *1*, No. 140022. (b) Ruddigkeit, L.; Van Deursen, R.; Blum, L. C.; Reymond, J.-L. Enumeration of 166 billion organic small molecules in the chemical universe database GDB-17. *J. Chem. Inf. Model.* **2012**, *52*, 2864−2875.

(17) Sun, J.; Wu, J.; Song, T.; Hu, L.; Shan, K.; Chen, G. Alternative approach to chemical accuracy: a neural networks-based first-principles method for heat of formation of molecules made of H, C, N, O, F, S, and Cl. *J. Phys. Chem. A* **2014**, *118*, 9120−9131.

(18) Fey, M.; Lenssen, J. E. *Fast graph representation learning with PyTorch Geometric*. 2019, arXiv preprint arXiv:1903.02428.

(19) Kingma, D. P.; Ba, J. *Adam: A method for stochastic optimization*. 2014, arXiv preprint arXiv:1412.6980.

(20) Efron, B. Bootstrap methods: another look at the jackknife. In *Breakthroughs in statistics*; Springer, 1992; pp 569−593.

(21) Schütt, K. T.; Kindermans, P.-J.; Sauceda, H. E.; Chmiela, S.; Tkatchenko, A.; Müller, K.-R. *Schnet: A continuous-filter convolutional neural network for modeling quantum interactions*. 2017, arXiv preprint arXiv:1706.08566.

(22) (a) Neese, F. The ORCA program system. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2012**, *2*, 73−78. (b) Neese, F. Software update: the ORCA program system, version 4.0. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2018**, *8*, No. e1327.

(23) (a) Cybenko, G. Approximation by superpositions of a sigmoidal function. *Math. Control Signals Syst.* **1992**, *5*, 455−455. (b) Hornik, K. Approximation capabilities of multilayer feedforward networks. *Neural Networks* **1991**, *4*, 251−257.

(24) Kennard, R. W.; Stone, L. A. Computer aided design of experiments. *Technometrics* **1969**, *11*, 137−148.

(25) Galvao, R. K. H.; Araujo, M. C. U.; José, G. E.; Pontes, M. J. C.; Silva, E. C.; Saldanha, T. C. B. A method for calibration and validation subset partitioning. *Talanta* **2005**, *67*, 736−740.

(26) Alder, B. J.; Wainwright, T. E. Studies in molecular dynamics. I. General method. *J. Chem. Phys.* **1959**, *31*, 459−466.

(27) Karplus, M. Development of multiscale models for complex chemical systems: from H+ H_2 to biomolecules (Nobel lecture). *Angew. Chem., Int. Ed.* **2014**, *53*, 9992−10005.

(28) Kmiecik, S.; Gront, D.; Kolinski, M.; Wieteska, L.; Dawid, A. E.; Kolinski, A. Coarse-grained protein models and their applications. *Chem. Rev.* **2016**, *116*, 7898−7936.

https://doi.org/10.1021/acs.jpca.2c02957
J. Phys. Chem. A 2022, 126, 6295−6300