Approaches for machine learning intermolecular interaction energies

Derek P. Metcalf
Center for Computational Molecular Science and Technology, School of Chemistry and Biochemistry
Georgia Institute of Technology
Atlanta, Georgia 30332-0400
`metcalf@gatech.edu`

Alexios Koutsoukas
Molecular Structure and Design
Bristol-Myers Squibb Company
P. O. Box 5400, Princeton, NJ 08543

Steven A. Spronk
Molecular Structure and Design
Bristol-Myers Squibb Company
P. O. Box 5400, Princeton, NJ 08543

Brian L. Claus
Molecular Structure and Design
Bristol-Myers Squibb Company
P. O. Box 5400, Princeton, NJ 08543

Deborah Loughney
Molecular Structure and Design
Bristol-Myers Squibb Company
P. O. Box 5400, Princeton, NJ 08543

Stephen R. Johnson
Molecular Structure and Design
Bristol-Myers Squibb Company
P. O. Box 5400, Princeton, NJ 08543

Daniel L. Cheney
Molecular Structure and Design
Bristol-Myers Squibb Company
P. O. Box 5400, Princeton, NJ 08543

C. David Sherrill
Center for Computational Molecular Science and Technology, School of Chemistry and Biochemistry
Georgia Institute of Technology
Atlanta, Georgia 30332-0400
`sherrill@gatech.edu`

Abstract

Accurate prediction of intermolecular interaction energies is a fundamental challenge in chemistry despite their pervasiveness in describing fundamental physical phenomena in pharmacology, biology, and materials science. Symmetry adapted perturbation theory (SAPT) provides rigorous quantum mechanical means for computing such quantities directly and accurately, but for a prohibitive computational cost in all but the smallest systems. We report accurate, low-cost supervised learning approaches for the prediction of interaction energies. Our work features data augmentation, specialized atomic descriptors, and the physically interpretable energy decomposition from SAPT as learning targets to address the idiosyncrasies of the intermolecular problem.

Second Workshop on Machine Learning and the Physical Sciences (NeurIPS 2019), Vancouver, Canada.

1 Introduction

1.1 Intermolecular interaction energy

Numerous phenomena in biology, pharmacology, and materials science can be explained by non-covalent interactions (NCIs). [7, 9, 14] High accuracy quantification of NCIs can be achieved using the conventional tools of quantum chemistry, including but not limited to coupled-cluster theory and perturbation theory.[2, 1] Symmetry adapted perturbation theory (SAPT)[15, 8] computes the interaction energy directly as a perturbation to the molecular systems with very high accuracy. Moreover, SAPT decomposes naturally into several energy contributions that can be used to characterize the nature of an interaction. For example, the simplest truncation of the SAPT expansion is dubbed SAPT0 and can be written as[10]

$$E_{SAPT0} = E_{elst} + E_{exch} + E_{ind} + E_{disp} \tag{1}$$

The terms of this truncation represent electrostatics, exchange, induction (or polarization), and London dispersion, respectively. Each term in this expression reflects an interpretable and physically meaningful contribution to the interaction energy[11] and are computed directly from quantum mechanics. Wavefunction methods like SAPT, while very accurately approximating the true interaction energy, become prohibitively expensive for large systems. This necessitates accurate low-cost approximations to address many interesting chemical problems.

1.2 Behler-Parrinello neural networks

In recent years, Behler-Parrinello neural networks (BPNNs)[3, 4] have become a quintessential tool for building models to describe potential energy and other properties in molecules in a maximally flexible way.[12] The BPNN relies on the separability of a molecular property into atomic contributions, where a feed-forward neural network infers only the atomic contributions to the total molecular property. Usually atomic contributions are combined by a simple sum, though more complicated schemes have been explored.[6]

Typically, BPNNs use a different neural network for each “atom-type,” that is, carbon, hydrogen, oxygen, etc. Each atom in a system is represented by an atomic environment vector, often so-called “symmetry functions” which encode the local environment of the atom in terms of radial and angular proximity with other atoms in the system. This architecture has the advantage of growing linearly in number of atom-types treated and learning transferable characteristics between atoms of the same identity. BPNNs also boast linear inference-time scaling in number of atoms in the system due to the atom-in-molecule scheme.[16]

2 Methods

In order to isolate factors affecting machine learned prediction of interaction energies, we study a model dataset consisting of 9000 configurations of NMe-acetamide / Aniline dimer. Interaction energy labels are obtained at the SAPT0 / jun-cc-pVDZ level of theory. Tests are performed on 47 crystallographic examples of the same dimer. We anticipate the findings from this model case will represent useful practices for multi-dimer machine learned potentials and potentials trained on reference data produced at higher levels of theory.

2.1 Data augmentation

Data that describes molecular properties is simply a collection of Cartesian coordinates and the identity of the nucleus at each coordinate. We study a representative constrained intermolecular version of this problem – training a neural network to the SAPT0 interaction energy of a single dimer system in a wide variety of conformations, rather than a diverse set of dimers.

In this work, we highlight an idiosyncrasy of data curation for the interaction energy case. One might be tempted to construct a training set for a single potential by scanning along many distances and Euler angles between two internally static monomers. This has the effect of capturing the most important features in the interaction energy surface like range-dependent attractions and anisotropy in energy during rotations of one monomer with respect to the other. Fixing the monomers to be

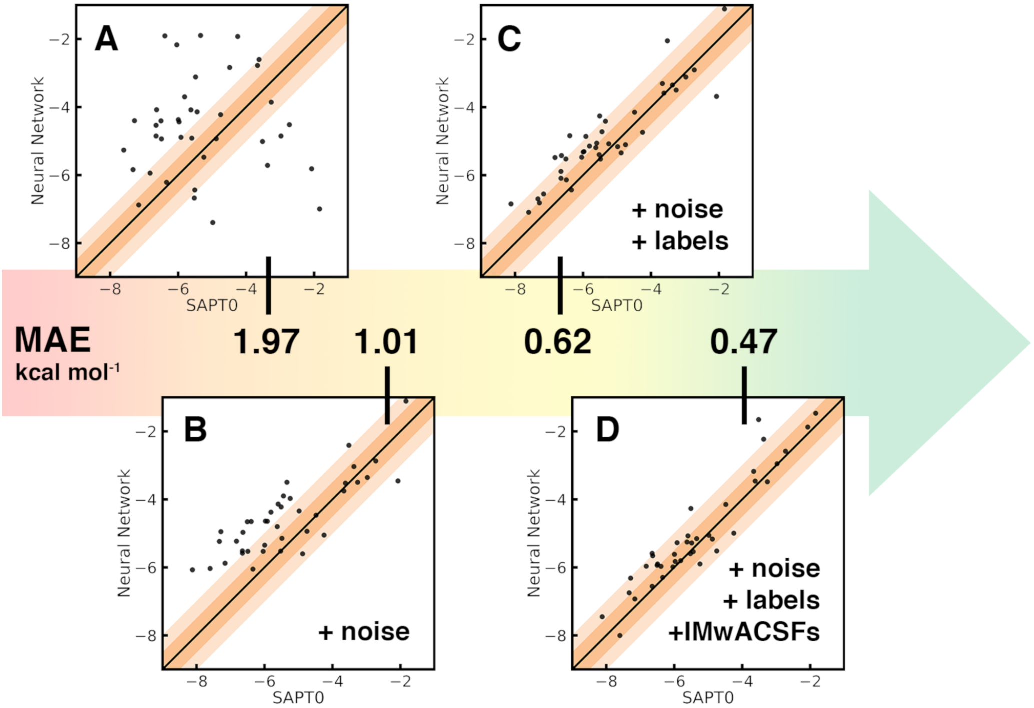

Figure 1: Four sequentially improved BPNN models for prediction of 47 crystallographic NMe-acetamide / Aniline dimer SAPT0 total interaction energies. All neural networks are trained on 9000 configurations. Shown is the SAPT0 target total interaction energy compared to the neural network-predicted total interaction energy. Dark orange corresponds to 0.5 kcal mol^{-1} from the target energy and light orange corresponds to within 1 kcal mol^{-1}. **(A)** was trained on only artificially generated configurations from Euler angles and distances with internally static monomers and represented with traditional wACSFs. **(B)** was trained with all configurations from A, but with all atomic Cartesian coordinates augmented by random perturbation between -0.1 and 0.1 Å. New energy labels are not provided, and all configurations use their mother energy label, consisting of the total interaction energy and its four SAPT0 components, weighted 60% and each 10%, respectively. **(C)** was trained on all of the configurations from B but with correct SAPT0 labels provided. **(D)** was trained with correctly labeled perturbed coordinates, but with the input descriptor represented as specialized intermolecular wACSFs (IMwACSFs).

internally static appears sensible because intuitively, the interaction energy varies little with respect to the possible small changes within the monomer. A neural network, however, lacks this intuition; symmetry function descriptors are inherently ordered by distance and some will change negligibly in a training set of internally static monomers. This effect causes any test sample with different internal monomer coordinates to have before-unseen descriptor values. The network is provided no information on how to adjust its prediction with respect to small internal changes, so they vary drastically and erroneously. We probe this effect in the intermolecular case for one dimer by training three data sampling techniques: one using only Euler angles and distances with internally static monomers, one using the same Euler angles and distances with randomly added Cartesian noise to every atomic coordinate (between -0.1 and 0.1 Å) without regenerating the correct SAPT0 energy, and lastly the same noisy Cartesian coordinates paired with the correct SAPT0 energies. Figure 1 parts A, B, and C illustrate this effect, notably drastically improved accuracy when noise is added to the Cartesian coordinates even when proper SAPT0 labels are not provided. This is a unique effect to the molecular case since both inter- and intramolecular degrees of freedom must be varied to capture even very weak dependencies on position. Molecular dynamics has been used to sample out-of-equilibrium configurations of molecular structures, which would adequately explore both degrees of freedom for interaction energies, but also may require ad hoc restrictions for keeping dimers bound in meaningful contact.[13]

2.2 Intermolecular Atomic Descriptors

Traditional Behler-Parrinello atom-centered symmetry functions (ACSFs)[4] and their descendants like the weighted atom-centered symmetry functions (wACSFs) of Marquetand and coworkers[5] provide reliable descriptions of local atomic environments while obeying the symmetries of a molecular system, such as translational and rotational invariances. The BPNN framework in conjunction with symmetry functions also accounts for invariance with respect to permutation of the same type of atom.

wACSFs have the form

$$G_i^{rad} = \sum_{j \neq i}^{N} Z_j e^{-\eta(r_{ij}-\mu)^2} f_c(r_{ij}) \tag{2}$$

$$G_i^{ang} = 2^{1-\zeta} \sum_{j,k \neq i}^{N} Z_j Z_k (1 + \lambda cos\theta_{ijk})^{\zeta} e^{-\eta(r_{ij}^2+r_{jk}^2+r_{ik}^2)} \times f_c(r_{ij}) f_c(r_{jk}) f_c(r_{ik}) \quad , \tag{3}$$

Each radial function G_i^{rad} for atom i has a unique η and μ hyperparameter pair, which correspond to Gaussian widths and shifts, resepectively, upon which other system atoms are evaluated. Similarly, angular functions G_i^{ang} depend on hyperparameters ζ, λ, and μ. All ACSF varieties assume some chemical locality, encoded in the cutoff function $f_c(r_{ij})$ which decays to 0 at a chosen cutoff radius.

Unlike the molecular problem, however, the intermolecular problem depends on the choice of which atoms belong to which molecule. Since there is no notion of molecule choice for molecular properties, there is a false symmetry in molecular descriptors like wACSFs that do not reflect this dependence. A false symmetry in the descriptor space is more harmful to model construction than a false asymmetry, since the latter can be rectified with sufficient data in a flexible model. As such, traditional molecular descriptors must be modified to address modeling intermolecular properties directly. A natural way to do this is to separate contributions to symmetry functions into same-molecule contributions and other-molecule contributions. Our test of this method on the NMA / Aniline model system is shown in Figure 1D and display notable generalization improvements.

2.3 Multi-target prediction

We leverage the shared information between SAPT components to recover both the physically meaningful component energies and the total interaction. We choose to train the neural networks to learn the collected electrostatics, exchange, induction, and dispersion energies. Each atom-type neural network has a densely connected final hidden layer to these energies, which are then constrained to sum to the total interaction energy. We choose our loss function to take the form

$$\mathcal{L} = (1-\gamma)\ MSE(\Delta E_{int}) + \gamma \sum_{i \in C} MSE(E_i) \tag{4}$$

with $C = \{\text{electrostatics, exchange, induction, dispersion}\}$. The parameter γ can be varied between 0 and 1, allowing for the loss function to include different proportions of component error and total interaction energy error. $\gamma = 0$ corresponds to the single-target training of total interaction energy and $\gamma = 1$ corresponds to equally weighting the fit to all component energies.

In Figure 2, we show that in our test system, scanning along the γ coordinate yields superior results to both $\gamma = 1$ and 0 by leveraging learning target relationships and encouraging systematic error cancellation of component predictions. Large γ do not benefit from cancellation of error, but low γ fails to recover the SAPT0 component energies. $\gamma = 0.6$ appears to recover very accurate component energies with improved total interaction energy accuracy compared to $\gamma = 0.0$.

3 Conclusions

We have introduced a framework and set of best practices for generating models for prediction of intermolecular interaction energies. These factors are general to any choice of statistical model that concedes to the atom-in-molecule prescription of BPNNs. These models, relying on only

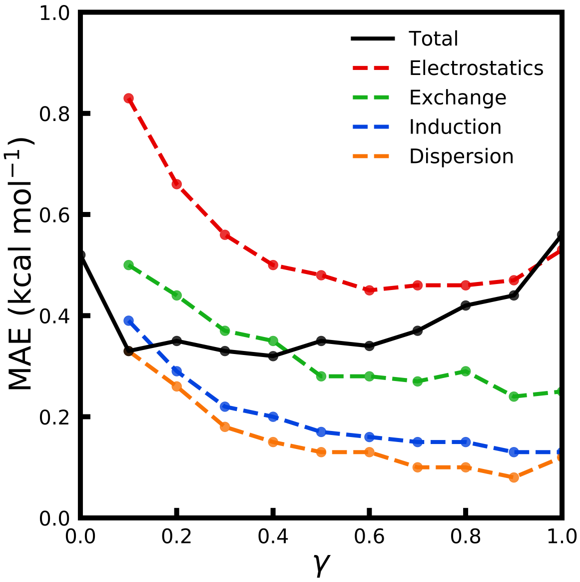

Figure 2: The validation errors of intermolecular BPNNs trained on 9000 configurations of N-Me-acetamide / Aniline dimer computed at the SAPT0 / jun-cc-pVDZ level of theory. γ is varied from 0.0 to 1.0, varying the loss function according to Eq. 4.

mathematically simple descriptors and neural network forward-passes can be evaluated in 130 μs atom^{-1}, orders of magnitude faster than the underlying SAPT which may take minutes or hours for dimers of interest.

Special care must be taken in data curation for NCIs where two sets of coordinates may be varied – the intramolecular coordinates, vital for capturing model behavior near geometric equilibria, and the intermolecular coordinates, capturing the behavior of NCIs as molecular positions vary with respect to one another. In the event full quantum mechanical reference data cannot be computed for new coordinates, data augmentation without explicitly relabeling along low-slope coordinates may be useful for more general models at no increased inference-time cost.

Intermolecular properties depend on monomer choice, so this dependence must be encoded into the feature space for use in statistical models. We provide one way of encoding this dependence, dubbed “intermolecular symmetry functions,” the underlying concept of which is extensible to other descriptors and architectures.

We leverage the relationship between SAPT components to multi-target predict both components and total interaction energy to high accuracy. This practice is useful anywhere a simple functional form synthesizes a desired property, but is especially valuable with SAPT, where component energies are independently physically meaningful in characterizing NCIs and simply sum to the total interaction energy.

In concert, these practices enable models in which diverse interacting systems can be accurately characterized with drastically reduced computational complexity compared to the quantum mechanical reference.

References

[1] R J Bartlett. Many-body perturbation theory and coupled cluster theory for electron correlation in molecules. *Annual Review of Physical Chemistry*, 32(1):359–401, 1981. doi: 10.1146/annurev.pc.32.100181.002043. URL `https://doi.org/10.1146/annurev.pc.32.100181.002043`.

[2] Rodney J. Bartlett and Monika Musiał. Coupled-cluster theory in quantum chemistry. *Rev. Mod. Phys.*, 79:291–352, Feb 2007. doi: 10.1103/RevModPhys.79.291. URL `https://link.aps.org/doi/10.1103/RevModPhys.79.291`.

[3] Jörg Behler and Michele Parrinello. Generalized neural-network representation of high-dimensional potential-energy surfaces. *Phys. Rev. Lett.*, 98:146401, Apr 2007. doi: 10.1103/PhysRevLett.98.146401. URL `https://link.aps.org/doi/10.1103/PhysRevLett.98.146401`.

[4] Jörg Behler. Atom-centered symmetry functions for constructing high-dimensional neural network potentials. *The Journal of Chemical Physics*, 134(7):074106, 2011. doi: 10.1063/1.3553717. URL `https://doi.org/10.1063/1.3553717`.

[5] M. Gastegger, L. Schwiedrzik, M. Bittermann, F. Berzsenyi, and P. Marquetand. wACSF—weighted atom-centered symmetry functions as descriptors in machine learning potentials. *The Journal of Chemical Physics*, 148(24):241709, 2018. doi: 10.1063/1.5019667. URL `https://doi.org/10.1063/1.5019667`.

[6] Justin Gilmer, Samuel S. Schoenholz, Patrick F. Riley, Oriol Vinyals, and George E. Dahl. Neural message passing for quantum chemistry. *ArXiv*, abs/1704.01212, 2017.

[7] M.M. Gromiha, K. Saraboji, S. Ahmad, M.N. Ponnuswamy, and M. Suwa. Role of non-covalent interactions for determining the folding rate of two-state proteins. *Biophys Chem. 2004 Feb 15;107(3):263-72. doi:*, 10., 2003.

[8] Bogumil Jeziorski, Robert Moszynski, and Krzysztof Szalewicz. Perturbation theory approach to intermolecular potential energy surfaces of van der waals complexes. *Chemical Reviews*, 94(7):1887–1930, 1994. doi: 10.1021/cr00031a008. URL `https://doi.org/10.1021/cr00031a008`.

[9] Elizabeth S. Lowe and Juan J.L. Lertora. Chapter 20 - dose–effect and concentration–effect analysis. In Arthur J. Atkinson, Shiew-Mei Huang, Juan J.L. Lertora, and Sanford P. Markey, editors, *Principles of Clinical Pharmacology (Third Edition)*, pages 343 – 356. Academic Press, third edition edition, 2012. ISBN 978-0-12-385471-1. doi: https://doi.org/10.1016/B978-0-12-385471-1.00020-9. URL `http://www.sciencedirect.com/science/article/pii/B9780123854711000209`.

[10] Trent M. Parker, Lori A. Burns, Robert M. Parrish, Alden G. Ryno, and C. David Sherrill. Levels of symmetry adapted perturbation theory (SAPT). I. Efficiency and performance for interaction energies. *The Journal of Chemical Physics*, 140(9):094106, 2014. doi: 10.1063/1.4867135. URL `https://doi.org/10.1063/1.4867135`.

[11] C. David Sherrill. Energy component analysis of interactions. *Accounts of Chemical Research*, 46(4):1020–1028, 2013. doi: 10.1021/ar3001124. URL `https://doi.org/10.1021/ar3001124`. PMID: 23020662.

[12] J. S. Smith, O. Isayev, and A. E. Roitberg. ANI-1: an extensible neural network potential with DFT accuracy at force field computational cost. *Chem. Sci.*, 8:3192–3203, 2017. doi: 10.1039/C6SC05720A. URL `http://dx.doi.org/10.1039/C6SC05720A`.

[13] Justin S. Smith, Ben Nebgen, Nicholas Lubbers, Olexandr Isayev, and Adrian E. Roitberg. Less is more: Sampling chemical space with active learning. , 148(24):241733, Jun 2018. doi: 10.1063/1.5023802.

[14] Christopher Sutton, Chad Risko, and Jean-Luc Brédas. Noncovalent intermolecular interactions in organic electronic materials: Implications for the molecular packing vs electronic properties of acenes. *Chemistry of Materials*, 28(1):3–16, 2016. doi: 10.1021/acs.chemmater.5b03266. URL `https://doi.org/10.1021/acs.chemmater.5b03266`.

[15] Krzysztof Szalewicz. Symmetry-adapted perturbation theory of intermolecular forces. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 2(2):254–272, 2012. doi: 10.1002/wcms.86. URL `https://onlinelibrary.wiley.com/doi/abs/10.1002/wcms.86`.

[16] Kun Yao, John E. Herr, David W. Toth, Ryker Mckintyre, and John Parkhill. The TensorMol-0.1 model chemistry: a neural network augmented with long-range physics. *Chem. Sci.*, 9:2261–2269, 2018. doi: 10.1039/C7SC04934J. URL `http://dx.doi.org/10.1039/C7SC04934J`.