Machine learning of interatomic potentials

Justin S. Smith
Metropolis Postdoctoral Fellow
Los Alamos National Lab
(T-CNLS/T-1)

Los Alamos National Lab:
Kipton Barros
Sergei Tretiak
Benjamin Nebgen
Nicholas Lubbers
Saryu Fensin
Timothy Germann

UNC Chapel Hill:
Roman Zubatyuk
Olexandr Isayev

University of Florida:
Adrian Roitberg

LA-UR-19-30185
Molecular (atomistic) dynamics

Energy

\[ E[r_1, r_2, ...] \]

Force

\[ f_i = -\nabla_i E \]

Dynamics

\[ m \frac{d^2 r_i}{dt^2} = f_i \]

In principle, requires a quantum mechanical calculation at each time step!
Ab initio Quantum Mechanics (QM)

Time-independent Schrödinger equation

\[ \hat{H} |\psi\rangle = E |\psi\rangle \]

\[ \hat{H} = KE + ER + EN + NN \]

Electrons treated as a probabilistic cloud

<table>
<thead>
<tr>
<th>Class</th>
<th>Method</th>
<th>Cost</th>
</tr>
</thead>
<tbody>
<tr>
<td>Semi-empirical</td>
<td>AM1, PM6, DFTB</td>
<td>O(N^2)</td>
</tr>
<tr>
<td>Density functional theory</td>
<td>B3LYP, wB97x, PBE</td>
<td>O(N^3)</td>
</tr>
<tr>
<td>Post-Hartree Fock</td>
<td>MP2, Coupled Cluster</td>
<td>O(N^4) &gt;</td>
</tr>
</tbody>
</table>
Classical force fields for Chemistry

\[ E_{FF} = \sum E_{ij}^{\text{nonbonded}} + \sum E_{ijkl}^{\text{Torsions}} + \sum E_{ijk}^{\text{Angles}} + \sum E_{ij}^{\text{Bonds}} \]

Classical Force Fields (FF):
CHARMM – OPLS – AMBER
Empirical potentials for materials

$E_i$ represents the energy of embedding atom $i$ within the electron field of its neighbors.

$$E = \sum_{i=0}^{N_{\text{atoms}}} F_\alpha \left( \sum_{i \neq j} \rho_\beta (r_{ij}) \right) + \frac{1}{2} \sum_{i \neq j} \phi_{\alpha\beta} (r_{ij})$$

Embedded Atom Model (EAM)

Levels of model chemistry

Classical $\mathcal{O}(n)$

Tight-binding $\mathcal{O}(n^2)$

DFT $\mathcal{O}(n) - \mathcal{O}(n^3)$

CCSD(T) $\mathcal{O}(n^7)$

Exact solution $\sim \exp(n)$

Machine learning!
Supervised Machine Learning

Types of tasks

• Regression
• Classification

A few applications

• Image recognition
• Social media moderation
• Stock market prediction

Training dataset for supervised learning

\[ \text{Inputs} = \{x_0, x_1, x_2, ..., x_N\} \]

\[ y_i = f^?(x_i) \]

\[ \text{Labels} = \{y_0, y_1, y_2, ..., y_N\} \]
Deep learning (neural networks) in a nutshell

\[ f(\vec{x} \cdot \vec{w}) \]

- **Activation function**
- **Pre-activation**
- **Weight vector**
- **Input vector**

- **Input layer**
- **Hidden layers**
- **Output layer**
Deep Learning for Atomistic Potentials

Training dataset for supervised learning

Inputs = \( \{ \mathbf{R}_0, \mathbf{R}_1, \mathbf{R}_2, \ldots, \mathbf{R}_N \} \)

\[ E_i = f^{QM}(\mathbf{r}_i) \]

Labels = \( \{ E_0, E_1, E_2, \ldots, E_N \} \)

Supervised Learning

Coordinates \( \mathbf{R} \)

Neural Network

Energies

Feedback

Prediction
What are deep learning potentials?

Molecular Fingerprints

Atomic Environment Descriptors

Property

DL Model

Atomic Property

Sum

DL Model

Atomic Property

Atomic Property

DL Model

DL Model

DL Model
Transferability and extensibility for DL potentials

Definition: **Transferability** – The ability for the model to predict on systems *not* in the training dataset

Definition: **Extensibility** – The ability for a model to predict on systems *larger* than those in the training dataset

Molecular fixed sized vector = no extensibility

Atomic fixed sized vector = extensibility
Building an energy conservative DL potential: how to get atomic forces?

Compute forces as $\vec{F} = -\nabla E$ if we enforce $E$ to be a smooth scalar field.

Input coordinates: $\vec{r}$
How do we represent the chemical environment of this oxygen?

Cutoff function:

\[
f_c(R_{ij}) = \begin{cases} 
0.5 \times \cos \left( \frac{\pi R_{ij}}{R_c} \right) + 0.5 & \text{for } R_{ij} \leq R_c \\
0.0 & \text{for } R_{ij} > R_c 
\end{cases}
\]
We choose physically inspired invariance:

Translation

Rotation

Permutation
ANI-1: an extensible neural network potential with DFT accuracy at force field computational cost†

J. S. Smith, O. Isayev, and A. E. Roitberg

Hierarchical modeling of molecular energies using a deep neural network

Nicholas Lubbers, Justin S. Smith, and Kipton Barros

Accurate and transferable multitask prediction of chemical properties with an atoms-in-molecules neural network

Roman Zubatyuk, Justin S. Smith, Jerzy Leszczynski, Olexandr Isayev
Common invariant components of ML model potentials

Radial symmetry functions


\[ G_{m}^{R} = \sum_{j \neq i} e^{-\eta(R_{ij}-R_{S})^2} f_{c}(R_{ij}) \]


\[ G_{m}^{R} = \sum_{j \neq i} e^{-(\xi_{ij}^{-1}+\eta^{-1})\gamma^2/2} f_{c}(R_{ij}) \]

Angular symmetry functions


\[ G_{m}^{A_{mod}} = 2^{1-\zeta} \sum_{j,k \neq i} (1 + \lambda \cos(\theta_{ijk}))^{\zeta} \exp \left[-\eta \left(R_{ij}^2 + R_{ik}^2 + R_{jk}^2 \right)^2 \right] f_{c}(R_{ij})f_{c}(R_{ik})f_{c}(R_{jk}) \]


\[ G_{m}^{A_{mod}} = 2^{1-\zeta} \sum_{j,k \neq i} (1 + \cos(\theta_{ijk} - \theta_{s}))^{\zeta} \exp \left[-\eta \left(\frac{R_{ij}^2 + R_{ik}^2}{2} - R_{S} \right) \right] f_{c}(R_{ij})f_{c}(R_{ik}) \]
ANI style neural network model potential

Radial environment
Distance over $j$ neighbors centered on atom $i$

Angular environment
Angle over $j,k$ neighbors centered on atom $i$

$$E = \sum_{i}^{N} E_i$$

Atomic Neural Network

Sum over $j$ neighbors

$Z_j$ and $Z_k$ type differentiation

Concatenate

HIP-NN style neural network model potential

Interaction is a message passed from neighboring atoms

\[ E_i = Q \cdot \sum_j E_j \]

Radial environment
Distance over \( j \) neighbors centered on atom \( i \)

\[ E = \sum_i E_i \]

One-hot Interaction over \( j \) neighbors

N Lubbers, JS Smith, K Barros; *J. Chem. Phys.*, 2018, 148, 241715
Can we predict when the model is wrong?

Ensemble disagreement can drive data generation

Good data coverage

Bad data coverage
Active Learning - The Big Picture
An automated and self-consistent data generation framework

Non-equilibrium Conformational sampler

Molecule Sampling (e.g. GDB small molecule database, small peptides, drug like molecules)

Structure Pools

Compute Cluster

Ensemble of ANI networks
Train network ensemble
ANI-1x Dataset (i.e. energies, forces, dipoles)

New test data
Check ensemble disagreement

Testing transferability and extensibility

ANI-MD Benchmark

128 frames from 1ns trajectories @ 300K for each:

DrugBank and Tripeptide Benchmarks

Various drug molecule

Chignolin

Trp-cage

CHNO

Extensibility

Count

Number of Atoms per Molecule

ANI-1x (Train + Test)

DrugBank

Tripeptides

Active-learning results vs. random sampling

**Dataset size comparison**

<table>
<thead>
<tr>
<th></th>
<th>ANI-1</th>
<th>ANI-1x</th>
</tr>
</thead>
<tbody>
<tr>
<td>Datapoints</td>
<td>22M</td>
<td>5M</td>
</tr>
</tbody>
</table>

**Relative E and F RMSE comparison**

![Graph showing relative energy (ΔE) and force (F) RMSE comparison between ANI-1 and ANI-1x across different datasets.]

**Active learning progression**

![Graph showing the decrease in RMSE with increasing training set size for ANI-1 and ANI-1x.]

*JS Smith, et al.; The Journal of Chemical Physics, (2018), 148 (24), 241733*
Molecular dynamics force errors vs. reference DFT
Correlation of force components for the ANI-MD Benchmark

Ranolazine
MAE=1.366
RMSE=1.943

Atazanavir
MAE=1.586
RMSE=2.414

Chignolin
MAE=1.972
RMSE=2.840

Trp-cage
MAE=2.099
RMSE=3.161

64 Atoms
103 Atoms
149 Atoms
312 Atoms
Transcriptional Regulatory Repressor Protein (5MXV) in explicit water
Simulated with ANI-2x (CHNOSFCl)

- ~35K atoms
- Explicit water
- No ions
- S, F and Cl in ligand

From Mycobacterium Tuberculosis

https://www.rcsb.org
Timings for a 5x ensemble prediction for ANI-2x

<table>
<thead>
<tr>
<th>GPU</th>
<th>ANI-2x time per step</th>
<th>Total time per step</th>
<th>Steps per day</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tesla V100</td>
<td>297ms</td>
<td>317ms</td>
<td>272k</td>
</tr>
</tbody>
</table>

0.5ns simulation time

RMSD (Å)
Simulation of Complex Chemical Reactions

Carbon nanoparticles/sheets nucleation [4000 atoms in 60A box at 2500K, 5ns MD simulation]
Transferring knowledge from DFT to CCSD(T)

- Subsample 10% of ANI-1x training data (0.5M of 5M)
- Recompute CCSD(T)/CBS level
- 340k parameters fixed, re-train 60k
- $10^7$ faster than DFT

Smith et al.; Nat. Comm. 2019
https://doi.org/10.1038/s41467-019-10827-4
Outsmarting Quantum Chemistry Through Transfer Learning
JS Smith, B Nebgen, R Zubatyuk, N Lubbers, C Devereux, K Barros, S Tretiak, O Isayev, A Roitberg
https://doi.org/10.1038/s41467-019-10827-4 Nat. Comm. 2019

- New ANI-1ccx model outperforms DFT on reaction energies and torsional profiles

- A 24 core hours calculation for CCSD(T)/CBS takes 2 GPU microseconds for ANI-1ccx

Torsion benchmark (CHNO only)
How to build a general potential for metal?

All possible configurations for a metal

Unphysical Configurations

Extreme Conditions

Liquid Phase

Crystal Phase

Where extreme conditions and rare events exist (e.g. shock simulations)

Typical sampling space for force fields and existing ML potentials

Recent published work

Linfeng Zhang, De-Ye Lin, Han Wang, Roberto Car, and Weinan E, Active Learning of Uniformly Accurate Inter-atomic Potentials for Materials Simulation, [arXiv:1810.11890]
How should we sample to build a general model?

All possible configurations for a metal

Can we learn the physics of crystals from sampling here?

LLNL Sierra (#2 on TOP500 list)

Thanks for the open science early access allocation!
Human vs machine driven sampling

Datasets:
**Machine** = MD simulation of **Random configurations** and active learning configurations
**Human** = MD melt simulation of **human selected crystal** with systematic configurations

**Machine@Human**: trained to **machine selected data** and tested on **human selected data**
Select crystal vs. random disorder MD sampling for Al

No human knowledge used in sampling

Sampling technique: Disorder only
(our current work)

Crystals chosen based on human knowledge (previous literature)

Linfeng Zhang, De-Ye Lin, Han Wang, Roberto Car, and Weinan E,
Active Learning of Uniformly Accurate Inter-atomic Potentials for Materials Simulation, [arXiv:1810.11890]
Energy, force, and stress errors for 25x 100ps MD simulations of liquid Al with temperatures in the range of 1000K to 2000K and volumes between 0.8 and 1.2 equilibrium.
Potential Energy

Dislocation Formation

- Potential composed of an ensemble of 2 Neural Networks potentials
- Simulations carried out on 80x Titan V GPUs in about 10 hours
- 2 kms$^{-1}$ shock
Thank you!