Accurate Energy Predictions via Machine Learning

Matthias Rupp
Fritz Haber Institute of the Max Planck Society, Berlin, Germany
Outline

1. Rationale
   ab initio simulations, machine learning

2. Kernel learning
   kernel trick, kernel ridge regression

3. Representations
   many-body tensor representation

4. Energy predictions
   energies of molecules and crystals
Rationale
Challenges in quantum mechanical simulations

High-throughput screening


Large systems


Long simulations


Quantum effects

Accurate simulations are limited by computational cost

<table>
<thead>
<tr>
<th>Method</th>
</tr>
</thead>
<tbody>
<tr>
<td>Full Configuration Interaction</td>
</tr>
<tr>
<td>Coupled Cluster</td>
</tr>
<tr>
<td>Møller-Plesset second order perturbation theory</td>
</tr>
<tr>
<td>Kohn-Sham Density Functional Theory</td>
</tr>
<tr>
<td>Tight Binding</td>
</tr>
<tr>
<td>Molecular Mechanics</td>
</tr>
</tbody>
</table>

Is it possible to be both accurate and fast?
In high-throughput settings, machine learning offers a new trade-off
Machine learning

Machine learning (ML) studies algorithms whose performance *improves with data* ("learning from experience").


Data $X \rightarrow$ Model $\hat{f}$

- No explicitly programmed problem-specific solution
- Systematic identification of regularity in data for prediction & analysis
- Interpolation in high-dimensional spaces
Interpolating electronic structure calculations with machine learning

Correlated calculations can be rapidly and accurately interpolated

- reference calculations, — ground truth, --- machine learning model

Assumes similar structure, similar property
Corresponding space or metric

Flexible functional form
Smoothness assumption

Physics in the examples
Flexibility versus physical constraints

Rupp, Int. J. Quant. Chem., 2015
## Relationship with other models

<table>
<thead>
<tr>
<th>ab initio</th>
<th>force fields</th>
<th>machine learning</th>
</tr>
</thead>
<tbody>
<tr>
<td>generally applicable</td>
<td>limited domain</td>
<td>generally applicable</td>
</tr>
<tr>
<td>no or little fitting</td>
<td>fitting to one class</td>
<td>refitted to any dataset</td>
</tr>
<tr>
<td>form from physics</td>
<td>form from physics</td>
<td>form from statistics</td>
</tr>
<tr>
<td>deductive</td>
<td>mostly deductive</td>
<td>inductive</td>
</tr>
<tr>
<td>few or no parameters</td>
<td>some parameters</td>
<td>many parameters</td>
</tr>
<tr>
<td>high accuracy</td>
<td>limited accuracy</td>
<td>high accuracy</td>
</tr>
<tr>
<td>slow</td>
<td>fast</td>
<td>in-between</td>
</tr>
<tr>
<td>small systems</td>
<td>large systems</td>
<td>large systems</td>
</tr>
</tbody>
</table>
Kernel learning
Learning with kernels

The kernel trick:

- Transform samples into higher-dimensional space
- Implicitly compute inner products there
- Rewrite linear algorithm using only inner products

\[ k: \mathcal{X} \times \mathcal{X} \to \mathbb{R}, \quad k(x, z) = \langle \Phi(x), \Phi(z) \rangle \]

Input space \( \mathcal{X} \)

Learning with kernels

The kernel trick:

- Transform samples into higher-dimensional space
- Implicitly compute inner products there
- Rewrite linear algorithm using only inner products

\[
k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}, \quad k(x, z) = \langle \phi(x), \phi(z) \rangle
\]

Kernel functions

Kernels correspond to **inner products**

Encode information about lengths and angles:

$$||x - z||^2 = \langle x, x \rangle - 2 \langle x, z \rangle + \langle z, z \rangle, \cos \theta = \frac{\langle x, z \rangle}{||x|| \cdot ||z||}$$

If $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ is symmetric positive semi-definite, then $k(x, z) = \langle \phi(x), \phi(z) \rangle$ for some $\phi : \mathcal{X} \to \mathcal{H}$

**Useful properties**

- Closed convex cone structure
- Natural interface $K_{ij} = k(x_i, x_j)$
- $\mathcal{X}$ can be any non-empty set
Examples of kernel functions

**Linear kernel**
\[ k(x, z) = \langle x, z \rangle \]

**Gaussian kernel**
\[ \exp(-\|x - z\|^2/2\sigma^2) \]

**Laplacian kernel**
\[ \exp(-\|x - z\|_1/\sigma) \]
Regression with kernels

**Linear ridge regression**

For models

$$f(x) = \sum_{i=1}^{d} \beta_i x_i$$

minimizing

$$\min_{\beta \in \mathbb{R}^d} \sum_{i=1}^{n} (f(x_i) - y_i)^2 + \lambda \|\beta\|^2$$

yields

$$\beta = (X^T X + \lambda I)^{-1} X^T y$$

---

**Kernel ridge regression**

For models

$$f(x) = \sum_{i=1}^{n} \alpha_i k(x_i, x)$$

minimizing

$$\min_{\alpha \in \mathbb{R}^n} \sum_{i=1}^{n} (f(x_i) - y_i)^2 + \lambda \|\alpha\|_H^2$$

yields

$$\alpha = (K + \lambda I)^{-1} y$$
A basis function picture of kernel regression

Weighted basis functions placed on training samples $x_i$

- learned $f(x)$
- basis functions $k(x_i, x)$
- prediction $\hat{f}(x) = \sum_i \alpha_i k(x_i, x)$

Vu et al., Int. J. Quant. Chem., 2015; Rupp, Int. J. Quant. Chem., 2015
The Gaussian process view

- Generalization of normal distribution
- Covariance = kernel
- Condition prior on data for posterior
- Variance as uncertainty estimate
Representations
Representations and Hilbert spaces

Numerical **representation** for arbitrary poly-atomic systems required
Corresponding Hilbert space as feature space

No unique solution
Example:
\[(a, b) \mapsto (a^2, \sqrt{2}ab, b^2)\]
\[(a, b) \mapsto (a^2, ab, ab, b^2)\]

Several state-of-the-art representations
symmetry functions
bispectrum, smooth overlap of atomic positions
moment tensor potentials
**many-body-tensor** and distribution-based representations
Hilbert spaces for arbitrary atomistic systems

Representation = numerical encoding of atomistic system for accurate interpolation

Requirements:

(i) **Invariance** against transformations preserving the property
translation, rotation, homonuclear permutations

(ii) **Uniqueness**: different in property ⇒ different in representation
reconstructability

(iii) **Smoothness**: continuous, ideally differentiable

(iv) **Generality**: work with any atomistic system
molecules, periodic systems

(v) **Efficiency**: fast to compute, require few reference calculations

(vi) **Simplicity**: conceptually straightforward

The many-body tensor representation

Distribution of $k$-body terms stratified by chemical elements:

$$f_k(x, z) = \sum_{i=1}^{N_a} w_k(i) D(x, g_k(i)) \prod_{j=1}^{k} C_{z_j, z_i}$$

[Huo & Rupp, arXiv, 2017]
Examples: molecules and bulk crystals
Example: surfaces

N atom on Pd(111) surface  MBTR of 2nd order (N-Pd)

top-sites  hollow-sites

Yasunobu Ando, work in progress, 2018
Energy predictions
The combinatorial nature of chemical and materials space

Molecule space
Graph theory

Materials space
Group theory

Combinatorial explosion
Machine learning

Left: aspirin derivatives
Learning potential energy surfaces

Figure: Chang & von Lilienfeld, CHIMIA, 2014
Enthalpies of formation of small organic molecules

7211 small organic molecules; C, N, O, S, Cl, saturated with H
Relaxed geometries, formation energies at DFT/PBE level of theory

<table>
<thead>
<tr>
<th>Representation</th>
<th>Kernel</th>
<th>E / kcal mol$^{-1}$</th>
<th>α / Å$^3$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>RMSE</td>
<td>MAE</td>
</tr>
<tr>
<td>CM [1]</td>
<td>Laplacian</td>
<td>4.76</td>
<td>3.47</td>
</tr>
<tr>
<td>BoB [2]</td>
<td>Laplacian</td>
<td>2.86</td>
<td>1.79</td>
</tr>
<tr>
<td>BAML [3]</td>
<td>Laplacian</td>
<td>2.54</td>
<td>1.15</td>
</tr>
<tr>
<td>SOAP [4]</td>
<td>REMatch</td>
<td>1.61</td>
<td>0.92</td>
</tr>
<tr>
<td>MBTR</td>
<td>Linear</td>
<td>1.14</td>
<td>0.74</td>
</tr>
<tr>
<td>MBTR</td>
<td>Gaussian</td>
<td><strong>0.97</strong></td>
<td><strong>0.60</strong></td>
</tr>
</tbody>
</table>

Potential energy surfaces of organic molecules

Ab initio molecular dynamics; DFT/PBE, Tkatchenko-Scheffler van der Waals forces

<table>
<thead>
<tr>
<th>Molecule</th>
<th>DTNN MAE</th>
<th>GDML Matérn MAE</th>
<th>MBTR linear RMSE</th>
<th>MAE</th>
<th>MBTR Gaussian RMSE</th>
<th>MAE [kcal/mol]</th>
</tr>
</thead>
<tbody>
<tr>
<td>benzene</td>
<td>0.04</td>
<td>0.07</td>
<td>0.06</td>
<td>0.05</td>
<td>0.05</td>
<td><strong>0.04</strong></td>
</tr>
<tr>
<td>uracil</td>
<td>–</td>
<td>0.11</td>
<td>0.14</td>
<td>0.10</td>
<td>0.06</td>
<td><strong>0.04</strong></td>
</tr>
<tr>
<td>naphthalene</td>
<td>–</td>
<td>0.12</td>
<td>0.15</td>
<td><strong>0.11</strong></td>
<td>0.15</td>
<td><strong>0.11</strong></td>
</tr>
<tr>
<td>aspirin</td>
<td>–</td>
<td>0.27</td>
<td>0.26</td>
<td><strong>0.18</strong></td>
<td>0.32</td>
<td>0.22</td>
</tr>
<tr>
<td>salicylic acid</td>
<td>0.50</td>
<td>0.12</td>
<td>0.17</td>
<td>0.12</td>
<td>0.11</td>
<td><strong>0.08</strong></td>
</tr>
<tr>
<td>malonaldehyde</td>
<td>0.19</td>
<td>0.16</td>
<td>0.28</td>
<td>0.21</td>
<td>0.13</td>
<td><strong>0.10</strong></td>
</tr>
<tr>
<td>ethanol</td>
<td>–</td>
<td>0.15</td>
<td>0.22</td>
<td>0.16</td>
<td>0.10</td>
<td><strong>0.07</strong></td>
</tr>
<tr>
<td>toluene</td>
<td>0.18</td>
<td>0.12</td>
<td>0.16</td>
<td>0.12</td>
<td>0.15</td>
<td><strong>0.11</strong></td>
</tr>
</tbody>
</table>

GDML = Gradient Domain Machine Learning, Chmiela et al., Sci. Adv., 2017
Different decorations in elpasolite crystals

Elpasolites $\text{ABC}_2\text{D}_6$
Density functional theory
Relaxed geometry
12 k materials, 12 elements
[Faber et al, Phys Rev Lett, 2016]

Prediction error 8 meV/atom (0.1% of range)
Convergence threshold 10 meV/atom
Learning curves

11k ABC$_2$D$_6$ structures
12 chemical elements
DFT/PBE level of theory

4611 ABC$_2$ structures
22 chemical elements
DFT/PBE level of theory

--- 1,2,3 chemical elements removed
Phase diagrams of Pt-group/transition metal binary alloys

Identification of low-energy compositions in convex hull

Active learning to exclude high-energy structures

Saves up to half the calculations

MAE 39 meV/atom

The NOMAD 2018 Kaggle challenge

Christopher Sutton (Fritz Haber Institute, Max Planck Society)

Al-Ga-In sesquioxides
DFT, 3 k structures

$\Delta$-learning
Geometry from Vegard’s rule
Enthalpy of formation and band gaps relaxed

Errors:
27 meV/atom (energy)
180 meV (band gap)

www.kaggle.com/c/nomad2018-predict-transparent-conductors
The NOMAD 2018 Kaggle challenge

Christopher Sutton (Fritz Haber Institute, Max Planck Society)

Vegard’s rule geometries

Relaxed geometries
Binary alloys

With Gus Hart, Chandramouli Nyshadham, Brayden Bekker (Brigham Young University)

- RMSE / meV/atom
- $10 \times 1\ k + 0.6\ k$ structures
- DFT formation enthalpy
- Vegard’s rule geometries
- fcc, bcc, hcp lattices
- all structures up to 8 atoms
Reliability

“It is not the estimate that matters so much, as the degree of confidence with the opinion”

N. Taleb, 2004

Uncertainty estimates

Extrapolation versus interpolation

Domain of applicability

Left: Logarithmized predictive variance versus signed error for energy predictions

Reliability

RMSE 380 meV, MAE 200 meV, $R^2 = 0.9916$
RMSE 12 meV/atom, MAE 6 meV/atom, $R^2 = 0.9854$
Interpretability

- Explain individual predictions
  Right for the right reason?

- Sensitivity analysis
  Local gradients explain variation, but not the “why”

- How to quantify interpretability?
  How to interpret basis set expansions?
Summary

Interpolation of ab initio simulations
Accurate and fast, via machine learning

Learning with kernels
Kernel trick: implicit transformations

Representations
Many-body tensor representation

Energy predictions
Molecules, crystals, surfaces
Outlook: towards crystal structure prediction

Sampling → Machine learning → Ab initio calculations

structure energy → structure energy

comprehensive fast accurate
Tutorial

DOI 10.1002/qua.24954

Special Issue

DOI 10.1063/1.5043213

Code

Efficient C++ implementation with Python bindings available as part of my qmmlpack package.
To appear this year.

Datasets

Public datasets of ab initio calculations for molecules, solids and liquids available.
qmml.org
nomad-coe.eu
Acknowledgements

Group

Current

Marcel Langer (PhD student)
Alex Goeßmann (PhD student)
Fabio Hernández-Hernandez (visitor)

Former

Haoyan Huo (visitor)
Lucas Deecke (student)
Yasunobu Ando (visitor)

Funding

[Ipam logo] [NOMAD logo]