Bayesian Error Estimation in Density Functional Theory

Karsten W. Jacobsen
Jens Jørgen Mortensen
Kristen Kaasbjerg
Søren L. Frederiksen
Jens K. Nørskov
CAMP, Dept. of Physics, DTU

James P. Sethna
LASSP, Cornell Univ.
Density Functional Theory

- Widely used in many different scientific areas
- “Predictive power” (LDA, GGA, exact exchange...)
- The reliability is usually estimated based on experience with practical calculations – try a few different functionals – no systematic theory
- The reliability varies a lot with type of system/type of property
Cohesive Energy versus Structural Energy Differences

Typical cohesive energy error ~0.2-0.4 eV
Small structural energy differences (10-100 meV) calculated with high reliability

(Skriver, 1985)

Our results: Ensemble error prediction
A Question (or two)

• Is it possible to replace the experience and “general wisdom” of the practitioners with a systematic approach?

• Can the reliability of a physical model (DFT-GGA) be estimated more systematically?
Standard Bayesian Approach to Model Determination/Selection

Bayes' theorem:

Model: M with parameters $\theta$, Data: D

$$P(\theta|DM) = \frac{P(D|\theta M)}{P(D|M)} P(\theta|M)$$

$$P(\theta|DM) \propto \exp \left( - \sum_i (y_i - y_i^0)^2 / 2 \sigma^2 \right)$$

Maximal probability: Best fit $\rightarrow$ Least squares

Prior probability

$$y^0 = \theta_0 + \theta_1 x$$

Model parameters

Data points

Noise in data

Data calculated with model parameters

Obtain ensemble of model parameters

y

x

$\theta$
Standard Bayesian: Fitting a Noisy Sine-Function

Fitting a noisy sine function with 6th order polynomial

100 data points

Low noise, but bad model (3rd order polynomial)

1000 data points (not shown)

Ensemble fluctuations too small to describe deviations.
Different Approach for “Bad” Models

\[ P(\theta | MD) \propto \exp \left( -C(\theta)/T \right), \quad C(\theta) = \sum_i (y_i - y_i^0)^2 \]

Effective temperature given by the best-fit cost function:

\[ T = \frac{2 C_0}{N_p} \]

Number of parameters

Low noise and bad model (3\textsuperscript{rd} order polynomial)

Previously used to develop and evaluate interatomic potentials

Frederiksen, Jacobsen, Brown, Sethna
PRL 93, 165501 (2004)
Exchange energy involves enhancement factor:

\[ E_x[n] = \int d\mathbf{r} \, n(\mathbf{r}) \epsilon_x^{\text{unif}}(n) F_x(s) \]

\[ s = \left( \frac{1}{24 \pi^2} \right)^{1/3} \frac{\left| \nabla n \right|}{n^{4/3}} \]

The model:
Expansion of enhancement factor:

\[ F_x(s) = \sum_{n=1}^{N_p} \theta_n F_n(s) = \sum_{n=1}^{N_p} \theta_n \left( \frac{s}{s+1} \right)^{2n-2} \]

Use PBE densities.
Fast (energies are linear in parameters)

PBE: \[ F_x(s) = 1 + \kappa - \frac{\kappa}{1 + \mu s^2/\kappa} \]
RPBE: \[ F_x(s) = 1 + \kappa \left( 1 - e^{-\mu s^2/\kappa} \right) \]

Perdew, Burke, and Ernzerhof, PRL 77, 3865 (1996).
The Database

Atomization energies for 20 molecules and cohesive energies for 12 solids (per atom)

Cost function: 
\[ C(\theta) = \sum_i \left( E_i^{\text{exp}} - E_i^{\text{calc}}(\theta) \right)^2 \]

Calculations performed with new real-space multigrid PAW code “gridPAW”
(http://www.camp.dtu.dk/campos)
(Jens Jørgen Mortensen)

Self-consistent PBE densities used for all calculations.

Because x-energy and therefore total energy is linear in parameters only a few coefficients need to be calculated.
Overfitting versus Model Flexibility

Many parameters give flexibility to model (many fluctuations possible). Too many parameters lead to overfitting.

Train:
Minimization of cost function on 20-1=19 molecules

\[ C(\theta) = \sum_i \left( E_i^{\text{exp}} - E_i^{\text{calc}}(\theta) \right)^2 \]

Test:
Evaluate cost function for the molecule which was left out. Average over which molecule is left out.

We use three parameters in the following.
Best-Fit Enhancement Factor

Minimization of cost function

\[ C(\theta) = \sum_i \left( E_i^{\text{exp}} - E_i^{\text{calc}}(\theta) \right)^2 \]

\[ F_x(s) = \sum_{n=1}^{N_p} \theta_n F_n(s) = \sum_{n=1}^{N_p} \theta_n \left( \frac{s}{s + 1} \right)^{2n-2} \]

\[ \theta_{\text{best fit}} = (1.00, 0.19, 1.90) \]

Units of eV

<table>
<thead>
<tr>
<th></th>
<th>LDA</th>
<th>PBE</th>
<th>RPBE</th>
<th>best-fit</th>
</tr>
</thead>
<tbody>
<tr>
<td>molecules:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>mean abs.</td>
<td>1.46</td>
<td>0.35</td>
<td>0.21</td>
<td>0.24</td>
</tr>
<tr>
<td>mean</td>
<td>1.38</td>
<td>0.28</td>
<td>-0.01</td>
<td>0.12</td>
</tr>
<tr>
<td>solids:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>mean abs.</td>
<td>1.35</td>
<td>0.16</td>
<td>0.40</td>
<td>0.27</td>
</tr>
<tr>
<td>mean</td>
<td>1.35</td>
<td>-0.09</td>
<td>-0.40</td>
<td>-0.24</td>
</tr>
<tr>
<td>all:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>mean abs.</td>
<td>1.42</td>
<td>0.28</td>
<td>0.28</td>
<td>0.25</td>
</tr>
<tr>
<td>mean</td>
<td>1.37</td>
<td>0.14</td>
<td>-0.16</td>
<td>-0.02</td>
</tr>
</tbody>
</table>

s=0 limit consistent with LDA!
Ensemble of Enhancement Factors

\[ P(\theta | MD) \propto \exp\left(-C(\theta)/T\right) \]

Sampling of probability distribution using Metropolis

For a harmonic cost function a direct evaluation is simpler:

\[ P(\theta | MD) \propto e^{-\frac{1}{2} \sum_i \lambda_i \alpha_i^2 / T} \]

Bayesian error estimation:

\[ \sigma_{\text{BEE}}(O) = \sqrt{\frac{1}{N} \sum_{\mu=1}^{N} \left( O(\theta^\mu) - O_{\text{best-fit}} \right)^2} \]

Essentially no additional computational cost (no self-consistency)
So, does it work?

For 46 out of 64 observables the error is within the Bayesian Error Estimate.
Comparing BEE with Actual Errors

- Atomization/cohesive energies
- Bond lengths
- Vib. freq. and bulk moduli

\[ \frac{(O_{\text{best-fit}} - O_{\text{exp}})}{\sigma_{\text{BEE}}(O)} \]
Example CO/Cu(100)

Experiment: Top site preferred
DFT-PBE: Bridge site preferred

Best fit plus BEE:

\[ E_{\text{top}} = -0.64 \pm 0.30 \text{ eV} \]
\[ E_{\text{bridge}} - E_{\text{top}} = -26 \pm 70 \text{ meV} \]

Also “CO/Pt(111) Puzzle”

Calc. says fcc:
\[ E_{\text{fcc}} - E_{\text{top}} = -0.13 \text{ eV} \]

Expt says top site
Cu bcc-fcc Structural Energy Difference

Best fit plus BEE:

\[ E_c = 3.3 \pm 0.5 \text{ eV} \]

\[ E_{\text{bcc}} - E_{\text{fcc}} = 35 \pm 4 \text{ meV} \]
Limitations

- Error estimates depend on model (GGA) and database (binding energies for molecules and solids)

- Problems if
  - Property not represented in database
  - Model unable to describe property at all
    - Example: van der Waals interactions
Conclusions and Outlook

● Systematic and computationally fast approach to estimating error bars on DFT-GGA calculations

● Predicted error bars on energies may vary by several orders of magnitudes – in agreement with expt.

● Possibility for tailoring functionals to specific purposes (RPBE for chemisorption)

● Can be generalized to other xc funcs., higher order derivatives, exact exchange (orbital func.), “stepping up the xc-ladder”...