Deciphering the Contents of Chemically-Trained Neural Networks into Physical Intuition

Eric Berquist

June 15th, 2017

Roadmap

- Quantum chemical calculations of EPR parameters aid in the interpretation and rationalization of experimental results
- DFT calculations of EPR g-tensors for Cu^{2+} systems show large deviations from experimental measurements and between different density functionals
- Understanding the differences in electronic structure between DFT and benchmark-quality wavefunction methods is necessary for proposing improvements to DFT
- Hypothesis: Parametrization of DFT is a viable method for improvement

Literature Survey

In published work, how do g-tensors calculated using density functional theory of Cu^{2+}-containing systems compare to experimental results?

Methodology of the Literature Analysis

For each paper:

- Extract available parameters: g_{xx}, g_{yy}, g_{zz}, g_{11}, g_{22}, g_{33}, $g_{\perp}$, $g_{\parallel}$, g_{iso}
- Where there are multiple calculations for one experimental data point, use the one with best agreement
- Calculate missing values of $g_{\parallel}$, g_{iso}, and $g_{\perp}$ if possible

For each of $g_{\parallel}$, g_{iso}, and $g_{\perp}$:

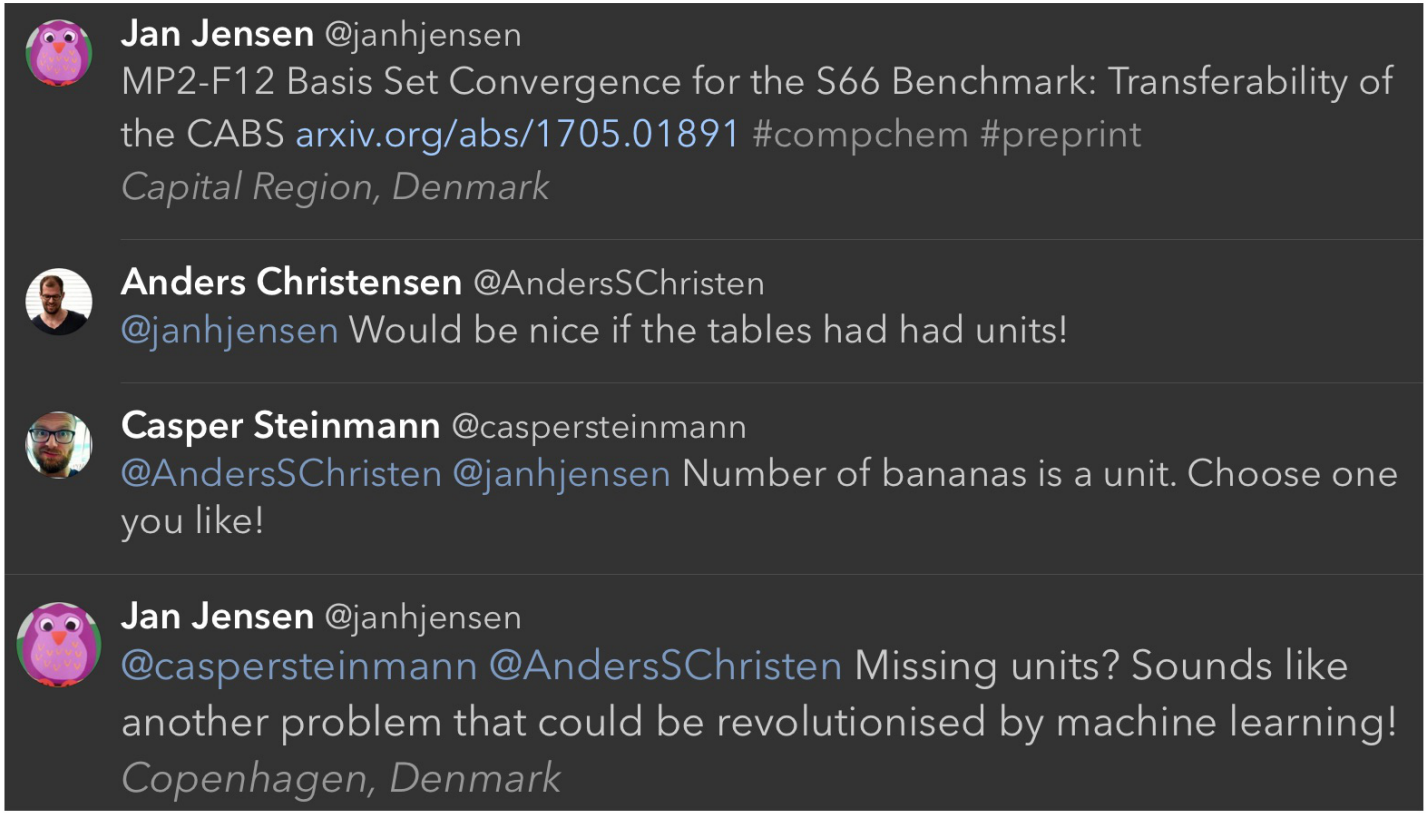
- Perform linear regression for the set of points
- Calculate the "percent incorrect trend"

What is machine learning

Arthur Samuel, 1959, the subfield of computer science that gives:

> *computers the ability to learn without being explicitly programmed.*
>
> *A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P if its performance at tasks in T, as measured by P, improves with experience E.*


THIS IS YOUR MACHINE LEARNING SYSTEM?
YUP! YOU POUR THE DATA INTO THIS BIG PILE OF LINEAR ALGEBRA, THEN COLLECT THE ANSWERS ON THE OTHER SIDE.
WHAT IF THE ANSWERS ARE WRONG?
JUST STIR THE PILE UNTIL THEY START LOOKING RIGHT.
DATA
ANSWERS


Jan Jensen @janhjensen
MP2-F12 Basis Set Convergence for the S66 Benchmark: Transferability of the CABS arxiv.org/abs/1705.01891 #compchem #preprint
Capital Region, Denmark
Anders Christensen @AndersSChristen
@janhjensen Would be nice if the tables had had units!
Casper Steinmann @caspersteinmann
@AndersSChristen @janhjensen Number of bananas is a unit. Choose one you like!
Jan Jensen @janhjensen
@caspersteinmann @AndersSChristen Missing units? Sounds like another problem that could be revolutionised by machine learning!
Copenhagen, Denmark

An (imperfect) analogy between neural networks and quantum chemistry

- The fundamental components of the network (kind of neuron activation functions, convolution or direct connection) are like the *Hamiltonian*.
- The number of components in each network layer and the number of layers are like the size of the *basis set*.

A better analogy

between relevance propagation and interaction energy decomposition approaches (SAPT, EDA):

- I am an item

Definitions of trained molecular properties

- Zero-point (vibrational) energy: $E_{\mathrm{ZPVE}} = \frac{1}{2} h \sum_i^{\text{normal modes}} \nu_i$
- Isotropic polarizability (static, $\omega = 0$):
  $\alpha_{\mathrm{iso}} = \bar{\alpha} \equiv \frac{1}{3}(\alpha_{xx} + \alpha_{yy} + \alpha_{zz})$
- Parallel 1st hyperpolarizability (static, $\omega_a = \omega_b = 0$): $\beta_{\parallel} \equiv$

Timeline

A future challenge: building databases

- GDB9/QM9 is the most commonly-used training set, the equivalent of the MNIST set of $\tilde{1}$0,000 labeled handwritten digits.
- It is now suffering from the same problem as MNIST: it is too simple and not representative of real-world training cases (molecules).
- Analogy: Pople basis sets (6-31G and derivatives) are still extremely common, not even because we don't know better, but because we "need to compare to past work".
- If a *general and transferable* ML model fails on GDB9, that is a warning sign, but the above cannot be a reason against extending deeper into chemical space for ML model training.

Backup Slides

Hyperpolarizability equations from paper

When the dipole moment coincides with the j-axis, we have

$$\beta_{\parallel} = \frac{3}{5}\beta_j = \frac{1}{5}\sum_{i=x,y,z}(\beta_{iij} + \beta_{iji} + \beta_{jii}), \tag{1}$$

or in the general case,

$$\beta_{\parallel} = \frac{3}{5|\mu|}\sum_{i,j=x,y,z}\beta_{iij}\mu_j, \tag{2}$$

where

$$\beta_{ijk} = \langle\langle \mu_i; \mu_j, \mu_k \rangle\rangle\,. \tag{3}$$