Journal of Cheminformatics

POSTER PRESENTATION Open Access

Modeling of molecular atomization energies using machine learning

Matthias Rupp[1*], Alexandre Tkatchenko[2], Klaus-Robert Müller[1], O Anatole von Lilienfeld[3]

From 7th German Conference on Chemoinformatics: 25 CIC-Workshop
Goslar, Germany. 6-8 November 2011

Atomization energies are an important measure of chemical stability. Machine learning is used to model atomization energies of a diverse set of organic molecules, based on nuclear charges and atomic positions only [1]. Our scheme maps the problem of solving the molecular time-independent Schrödinger equation onto a non-linear statistical regression problem. Kernel ridge regression [2] models are trained on and compared to reference atomization energies computed using density functional theory (PBE0 [3] approximation to Kohn-Sham level of theory [4,5]). We use a diagonalized matrix representation of molecules based on the inter-nuclear Coulomb repulsion operator in conjunction with a Gaussian kernel. Validation on a set of over 7000 small organic molecules from the GDB database [6] yields mean absolute error of ~10 kcal/mol, while reducing computational effort by several orders of magnitude. Applicability is demonstrated for prediction of binding energy curves using augmentation samples based on physical limits.

Author details
[1]Machine Learning Group, Technical University, Berlin, 10587, Germany. [2]Fritz-Haber-Institute, Max-Planck Society, Berlin, 14195, Germany. [3]Argonne National Laboratory, Argonne, Illinois, 60439, USA.

Published: 1 May 2012

References

1. Rupp M, Tkatchenko A, Müller KR, von Lilienfeld OA: **Fast and Accurate Modeling of Molecular Atomization Energies with Machine Learning.** *Physical Review Letters* 2012, **108(5)**.
2. Hastie T, Tibshirani R, Friedman J: **The Elements of Statistical Learning: Data Mining, Inference, and Prediction.** New York: Springer; 2009.
3. Perdew JP, Ernzerhof M, Burke K: **Rationale for mixing exact exchange with density functional approximations.** *J Phys Chem* 1996, **105(22)**:9982-9985.
4. Hohenberg P, Kohn W: **Inhomogeneous Electron Gas.** *Phys Rev* 1964, **136(3B)**:B864-B871.
5. Kohn W, Sham LJ: **Self-Consistent Equations Including Exchange and Correlation Effects.** *Phys Rev* 1965, **140(4A)**:A1133-A1138.
6. Blum LC, Reymond JL: **970 Million Druglike Small Molecules for Virtual Screening in the Chemical Universe Database GDB-13.** *J Am Chem Soc* 2009, **131(25)**:8732-8733.

doi:10.1186/1758-2946-4-S1-P33
Cite this article as: Rupp *et al.*: Modeling of molecular atomization energies using machine learning. *Journal of Cheminformatics* 2012 **4** (Suppl 1):P33.

* Correspondence: matthias.rupp@tu-berlin.de
[1]Machine Learning Group, Technical University, Berlin, 10587, Germany
Full list of author information is available at the end of the article