

POSTER PRESENTATION

Open Access

Kernel density estimation of CSD distributions - an application to knowledge based molecular optimisation

Patrick McCabe^{1*}, Oliver Korb¹, Jason Cole¹, Robin Taylor²

From 9th German Conference on Chemoinformatics
Fulda, Germany. 10-12 November 2013

The Cambridge Structural Database (CSD) contains a large amount of molecular structure data (bond length, bond angle and torsion angle data.) Much of this data has previously been extracted in histogram form and provided in the Mogul program. Histograms however have several disadvantages e.g. they are not smooth, they depend on bin widths and bin end points.

Kernel density estimators do not bin data and have no end points but centre a kernel function at each data point and smooth kernel functions will generate smooth density estimates [1]. A difficulty of the approach though is how wide to make the kernel functions.

In this work kernel density estimation is used to generate probability density functions (pdfs) for bond length, bond angle and torsion angle histograms derived from the CSD. Gaussian kernels are used for bond length and bond angle data and a von Mises kernel is used for the torsion angle data [2]. The resulting pdfs are smooth and are suitable for application to molecular geometry optimisation.

Authors' details

¹CCDC, 12 Union Road, Cambridge, CB2 1EZ, UK. ²Taylor Cheminformatics Software, 54 Sherfield Avenue, Rickmansworth, Herts., WD3 1NL, UK

Published: 11 March 2014

References

1. Silverman BW: *Density Estimation for Statistics and Data Analysis*. Chapman & Hall/CRC Monographs on Statistics & Applied Probability Chapman and Hall/CRC; 1986.
2. Evans M, Hastings NAJ & Peacock: *Statistical distributions*. Wiley; 2000.

* Correspondence: mccabe@ccdc.cam.ac.uk

¹CCDC, 12 Union Road, Cambridge, CB2 1EZ, UK

Full list of author information is available at the end of the article

doi:10.1186/1758-2946-6-S1-P10

Cite this article as: McCabe *et al.*: Kernel density estimation of CSD distributions - an application to knowledge based molecular optimisation. *Journal of Cheminformatics* 2014 **6**(Suppl 1):P10.

Publish with **ChemistryCentral** and every scientist can read your work free of charge

"Open access provides opportunities to our colleagues in other parts of the globe, by allowing anyone to view the content free of charge."

W. Jeffery Hurst, The Hershey Company.

- available free of charge to the entire scientific community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:
<http://www.chemistrycentral.com/manuscript/>



ChemistryCentral