

Oral presentation

Open Access

Beyond descriptor vectors: QSAR modelling using structural similarity

Andreas Zell*, G Hinselmann, NH Fechner and A Jahn

Address: University of Tuebingen, Sand 1, 72076 Tuebingen, Germany

* Corresponding author

from 3rd German Conference on Chemoinformatics
Goslar, Germany. 11-13 November 2007

Published: 26 March 2008

Chemistry Central Journal 2008, 2(Suppl 1):S3 doi:10.1186/1752-153X-2-S1-S3

This abstract is available from: <http://www.journal.chemistrycentral.com/content/2/S1/S3>

© 2008 Zell et al.

Kernel based machine learning methods like *support vector machines* or *gaussian processes* have gained increasing attention for QSAR modelling in recent years. One of the most interesting aspects of this method is the analogy between the kernel and a similarity measure. Each similarity measure that fulfils the kernel properties can be used as a kernel. But despite the possibility to incorporate structural/topological information directly into the similarity score, as it is done by state-of-the-art methods like *feature tree s[1]*, most studies that use kernel methods are limited to classical descriptor representations [2].

In this work we introduce some structural kernels that overcome these limitations and compare them to QSAR models based numerical encoding. The structural kernels regard the molecular topology as based on atom environments (*assignment kernels* [3]) or as based on adapted fingerprint representations. Additionally we compare the structural kernels to numerical kernels calculated for descriptor representation using a set of 200 molecular descriptors.

The structural and numerical kernels were used for training models with both *support vector machines* and *gaussian processes* on more than ten different QSAR datasets taken from BindingDB [4].

The results clearly indicate that in general the structural kernels are superior to the numerical encodings although no single kernel dominates. In the case that no further information is available it seems recommendable to use

the *optimal assignment kernel*[3] due to its overall good performance.

References

1. Rarey M, Dixon JS: *J Comp Aid Mol Design* 1998, **12**:471-490.
2. Schwaighofer A, et al.: *J Chem Inf Mod* 2007:401-424.
3. Fröhlich H, Wegner JK, Zell A: *Proc Int Joint Conf Neur Net (IJCNN)* 2005:913-918.
4. Chen X, Lin Y, Liu M, Gilson M: *Bioinformatics* 2002, **18**:130-139.