Chemistry Central Journal

Poster presentation

An extension of the pharmacophore kernel using radial atomtype fingerprints

G Hinselmann*, M Eckert, T Holder, A Jahn, N Fechner and A Zell

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

* Corresponding author

This abstract is available from: http://www.journal.chemistrycentral.com/content/3/S1/P11

The prediction of the biological activity of a chemical compound is a challenging task in Computational Chemistry and was restricted to vectorial representations of the molecular graph for decades. Kernel functions are positive semidefinite similarity measures that can be defined on arbitrary structured data. This class of similarity functions can be used in kernel-based machine learning algorithms. Interestingly, many graph kernel approaches from Computer Science share properties of traditional similarity measures for chemical compounds, like molecular fingerprints based on paths, cycles and subgraphs.

In this work, we present a hybrid technique derived from the Pharmacophore Kernel [1] and Radial Atom Environments [2]. Whereas in the original work [1] atoms were represented by their element number and partial charge, we employ a bounded depth-first-search to enumerate the complete neighbourhood of each non-carbon atom up to specified depth. Therefore, a pharmacophore can be defined by a triangle between three radial fingerprints. This opens the possibility to replace the simple Delta Kernel, which was used for the comparison of the vertices of two pharmacophores in the original work of Mahé *et al.* [1], by more sophisticated kernel functions defined on sets of discrete features. We tried the Tanimoto, MinMax [3] and the Spectrum Kernel [4] in our experiments. This results in valid kernel functions again, because a valid kernel function is replaced by another.

In the results section, we benchmark our approach against different state-of-the art graph kernels and the Radial Basis Function Kernel using descriptors calculated with drag-onX 1.4 on various QSAR data sets taken from the literature. The models were trained using the machine learning library LIBSVM. The results show that our approach improves the predictive power significantly on many benchmark problems.

References

1. Mahé P, et al.: J Chem Inf Mod 2006, 46:2003-2014.
2. Bender A, et al.: J Chem Inf Comp Sci 2004, 44:170-178.
3. Ralaivola L, et al.: Neural Networks Volume 18. Elsevier Science Ltd.; 2005:1093-1110.
4. Leslie C, et al.: Pacific symposium on biocomputing 2002.