

Poster presentation

Open Access

Estimating the applicability domain of kernel based QSPR models using classical descriptor vectors

NH Fechner*, G Hinselmann, C Schmiedl and Andreas Zell

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):P2 doi:10.1186/1752-153X-2-S1-P2

This abstract is available from: <http://www.journal.chemistrycentral.com/content/2/S1/P2>

© 2008 Fechner et al.

The propagation of machine learning based property prediction methods (e.g. QSAR, QSPR,...) has lead to the question of the reliability of the prediction. This leads to the development of methods enabling the estimation of the reliability of a model based prediction.

There are two principal approaches in dealing with this demand: estimating the expected derivation from the prediction (e.g. *gaussian processes*) or classifying each compound whether the model is specified for it or not. The last approach has become known as estimating the *applicability domain* [1,2] of a model. One drawback of the different AD estimation methods is that most of them are based on the spatial embedding of the training dataset in the descriptor space. Thus these algorithms are not directly suited in modelling the *applicability domain* of kernel-based predictors, which are working in a extremely high dimensional implicit feature space.

In this study we examined to what extent a standard descriptor based AD model can be used to describe the *applicability domain* of an *optimal assignment kernel* [3] based predictor. We split the popular Huuskonen [4] logS dataset 2:1 in a training and a test set and compared some standard AD methods [1,2] (range-based, convex hull, leverage,...) regarding the correlation of the estimated AD with the test error. The results indicate that it is possible to estimate the *applicability domain* of a kernel based model using classical descriptor encodings of the molecules. Furthermore the results show that there are significant differences between the different methods. In our application the geometrical convex hull approach was superior.

References

1. Jaworska J, Nikolova-Jeliazkova N, Aldenberg T: *ATLA* 2005:445-459.
2. Netzeva TI, Worth AP, Aldenberg T: *ATLA* 2005:1-19.
3. Fröhlich H, Wegner JK, Zell A: *Proc Int Joint Conf Neur Net (IJCNN)* 2005:913-918.
4. Huuskonen J: *J Chem Inf Mod* 2000:773-777.