

Poster presentation

Open Access

Additive inductive learning in QSAR/QSPR studies and molecular modeling

II Baskin*, NI Zhokhova, VA Palyulin and NS Zefirov

Address: Department of Chemistry, Moscow State University, 119991, Moscow, Russia

* Corresponding author

from 4th German Conference on Chemoinformatics
Goslar, Germany. 9–11 November 2008

Published: 5 June 2009

Chemistry Central Journal 2009, 3(Suppl 1):P25 doi:10.1186/1752-153X-3-S1-P25

This abstract is available from: <http://www.journal.chemistrycentral.com/content/3/S1/P25>

© 2009 Baskin et al; licensee BioMed Central Ltd.

In an ordinary inductive learning used in QSAR/QSPR studies a property y is approximated as a function of some fixed number of molecular descriptors:

$$y = f(x_1, \dots, x_m; c_1, \dots, c_p),$$

where x_i is the value of the i -th descriptor, c_j – is the values of the j -th learnable (adjustable) parameter, f – a non-linear function. Parameters c_j are learned using some training set of compounds.

We define the additive inductive learning (AIL) as a process, in which a property y of a chemical object (such as chemical compound) consisting of s subobjects (such as atoms, bonds, fragments, etc) is approximated in an additive manner:

$$y = \sum_{i=1}^s f(z_{i1}, \dots, z_{im}; c_1, \dots, c_p),$$

where z_{ij} is the value of the j -th descriptor for the i -th subobject. This definition can be extended to include several types of subobjects.

Due to the existence of many additive properties in chemistry and physics (the energy is the most prominent one), AIL is especially well suited for solving numerous problems in chemo- and bioinformatics, ranging from QSAR/QSPR studies up to molecular modeling, computing binding energies, etc.

We suggest two basic approaches to AIL: (i) the use of special additive neural networks (developed by us for this purpose), and (ii) the use of kernel-based approaches, such as the support vector machines, with a special type of kernels:

$$K_{ab} = \sum_i \sum_j k(a, i, b, j)$$

where K_{ab} is the kernel for the pair of objects a and b , while $k(a, i, b, j)$ is the "microkernel" for the pair of the i -th subobject in a and the j -th subobject in b .

Several examples of using AIL in chemoinformatics and the prospects of its application in molecular modeling are considered.