Machine Learning Methods in Property Prediction: *Quo Vadis?*

Igor I. Baskin[1], Alexandre Varnek[2]

[1]*Faculty of Physics, Lomonosov Moscow State University, 119991, Moscow, Russia*
[2]*Laboratoire de Chémoinformatique, UMR 7140 CNRS, Université de Strasbourg, 1, rue B. Pascal, 67000, Strasbourg, France*

Machine learning methods play very important role in chemoinformatics [1], especially for property prediction [2]. In this lecture, they are characterized in terms of the "modes of statistical inference" and "modeling levels" nomenclature and by considering different facets of the modeling with respect to input/ouput matching, data types, models duality, and models inference. Particular attention is paid to new approaches and concepts that may provide efficient solutions of common problems in chemoinformatics related to property prediction: inductive knowledge transfer (the use of multi-task learning and feature nets for improving property prediction) [3], one-class classification (its use in conducting virtual screening and defining applicability domain) [4], transductive learning (for dealing with small and unbalances datasets) [5], ensemble learning (the use of bagging, boosting and stacking for improving property prediction), the use of convolution kernels (for predicting properties without the use of global molecular descriptors) and the concept of continuous molecular fields (for building 3D QSAR models and conducting virtual screening based on the similarity of molecular fields) [6], etc.

[1] A.Varnek; I.I.Baskin. Mol. Inf. 30 (2011) 20-32.
[2] A.Varnek; I.Baskin. J. Chem. Inf. Model. 52 (2012) 1413-1437.
[3] A.Varnek; C.Gaudin; G.Marcou; I.Baskin; A.K.Pandey; I.V.Tetko. J. Chem. Inf. Model. 49 (2009) 133-144.
[4] I.Baskin; N.Kireeva; A.Varnek. Mol. Inf. 29 (2010) 581-587.
[5] E.Kondratovich, I.I.Baskin, A.Varnek. Mol. Inf. 32 (2013) 261-266.
[6] I.I.Baskin; N.I.Zhokhova. J. Comput. Aided. Mol. Des. 27 (2013) 427-442.