[P10] Reaction data treatment by means of Condensed graph of reaction

Gimadiev T.R.[1,2], Madzhidov T.I.[1], Nugmanov R.I. [1], Klimchuk O. [2], Casciuc I.V.[2], Baskin I.I.[3], Antipin I.S. [1], Varnek A.A.[1,2]

[1]*Kazan Federal University, Kazan, Russia,*
[2]*Strasbourg University, Strasbourg, France,*
[3]*Lomonosov Moscow State University, Moscow, Russia*

In this work we report new approach and best practices for curation of chemical reactions data applied to particular case of bimolecular nucleophilic substitution (S_N2) reactions. Some 8000 raw entities of S_N2 reactions was collected from the literature [1], including information about reaction rate (log*k*), temperature and solvent. For the curation of the dataset, the Condensed Graph of Reaction (CGR) approach [2] was used. The latter allows to encode all reactants and products in one sole molecular graph described by both conventional bonds (single, double, etc.) and dynamic bonds (single-to-double, broken single, created double, etc.) characterizing chemical transformations. In turn, this graph can be encoded by "CGR signature" which represent linear string describing the entire reaction or reaction center only. In such a way one can easily compare graphs, without using time-consuming graph embedding approach. This feature helps to group similar reactions. Reaction signatures could be used for identification of wrong atom-to-atom mapping.

Predictive model for log*k* has been built using Support Vector Regression and involved both ISIDA fragment descriptors [3] and physico-chemical parameters describing experimental conditions [4, 5]. Only structurally unique reactions were selected in the modeling set. The model displayed a reasonable performance in 5 fold cross validation: R^2=0.75 and RMSE=0.61 log*k* units. The latter is similar to intra-laboratory experimental error estimated as 0.5 log*k* units. External validation on a dataset of 93 reactions extracted from recent publications lead to reasonable statistics: RMSE=0.8 and R^2=0.64.

The research was supported by Russian Science Foundation, grant 14-43-00024.

Bibliography:

1. Palm VA (1977) Fundamentals of the Quantitative Theory of Organic Reactions, in Russian. Khimiya, Leningrad
2. Hoonakker F, Lachiche N, Varnek A, Wagner A (2011) Condensed Graph of Reaction: considering a chemical reaction as one single pseudo molecule. Int J Artif Intell Tools 20:253–270.
3. Varnek A, Fourches D, Horvath D, et al (2008) ISIDA - Platform for Virtual Screening Based on Fragment and Pharmacophoric Descriptors. Curr Comput Aided-Drug Des 4:191–198. doi: 10.2174/157340908785747465
4. Madzhidov TI, Polishchuk PG, Nugmanov RI, et al (2014) Structure-reactivity relationships in terms of the condensed graphs of reactions. Russ J Org Chem 50:459–463.
5. Gimadiev TR, Madzhidov TI, Nugmanov RI, et al (2018) Assessment of tautomer distribution using the condensed reaction graph approach. J Comput Aided Mol Des. doi: 10.1007/s10822-018-0101-6