MOLECULAR DESCRIPTORS BASED ON AUTOMORPHISM DATA

Kurt Varmuza,[a,b] Matthias Dehmer,[c,d] and Peter Filzmoser[a]

[a] *Vienna University of Technology, Institute of Statistics and Mathematical Methods in Economics, Computational Statistics, Wiedner Hauptstrasse 7/105-6, A-1040 Vienna, Austria; kurt.varmuza@tuwien.ac.at; peter.filzmoser@tuwien.ac.at*

[b] *Vienna University of Technology, Institute of Chemical, Environmental and Bioscience Engineering, Getreidemarkt 9/166-2, A-1060 Vienna, Austria*

[c] *Swiss Distance University of Applied Sciences, Department of Computer Science, Schinerstrasse 18, CH-3900 Brig, Switzerland; matthias.dehmer@ffhs.ch*

[d] *UMIT TIROL - Private University For Health Sciences and Health Technology, Eduard Wallnöfer Zentrum 1, A-6060 Hall in Tyrol, Austria*

Chemical structures are represented by colored mathematical graphs with vertices (atoms) defined by one of three elements (C, N, O), and edges (bonds) defined by one of four bond types (single, double, triple, aromatic). The complete automorphism group [1] of such chemical structure graphs is determined [2,3] and then evaluated by functions in the programming environment R [4].

Molecular descriptors based on automorphism data comprise symmetry and entropy measures, as well as roots of graph polynomials. For instance, the orbit polynomial [5] has been defined by orbit data for vertices and edges obtained from the automorphism data of graphs. In a previous work this concept was applied to exhaustive sets of isomeric alkanes [6] and is now extended to coloured graphs (CHNO molecules). Such descriptors have been used here together with already common ones for the creation of models for quantitative structure property relationships and for the prediction of chemical class memberships. Multivariate calibration models have been obtained by linear PLS regression optimized and evaluated by the strategy repeated double cross validation [7,8].

References:

[1] N. Trinajstic: *Chemical graph theory*, CRC Press, Boca Raton, FL, USA, 1992.
[2] K. Varmuza, et al.: *Croatica Chemica Acta*, **78** (2005) 141-149.
[3] H. Scsibrany, K. Varmuza: Software SubMat, 2004, http://www.lcm.tuwien.ac.at.
[4] R, A language and environment for statistical computing, R Development Core Team, Foundation for Statistical Computing, http://www.r-project.org, Vienna, Austria, 2024.
[5] M. Dehmer, et al.: *IEEE Access*, **8** (2020) 36100-36112.
[6] K. Varmuza, et al.: *Croatica Chemica Acta*, **94** (2021) 47-58.
[7] P. Filzmoser, B. Liebmann, K. Varmuza, *J. Chemometr.,* **23** (2009) 160-171.
[8] K. Varmuza, P. Filzmoser: *Introduction to multivariate statistical analysis in chemometrics*, CRC Press, Boca Raton, FL, USA, 2009.

Abstract for poster presentation at Conference Math/Chem/Com, 3-7 June 2024, Dubrovnik, Croatia