Similarity metrics under the magnifying glass: comparative study on metabolomic fingerprints

Anita Rácz[1], Filip Andrić[2], Dávid Bajusz[3], Károly Héberger[1]

[1]Plasma Chemistry Research Group, Research Centre for Natural Sciences, Hungarian Academy of Sciences, Hungary

[2]Faculty of Chemistry, University of Belgrade, Belgrade, Serbia

[3]Medicinal Chemistry Research Group, Research Centre for Natural Sciences, Hungarian Academy of Sciences, Hungary

Contemporary metabolomic fingerprinting is based on multiple spectrometric and chromatographic signals, used either alone or combined with structural and chemical information of metabolic markers at the semiqualitative level. However, signal shifting, convolution, and matrix effects may compromise metabolomic patterns. Recent increase in the use of qualitative metabolomic data, described by the presence (1) or absence (0) of particular metabolites, demonstrates great potential in the field of metabolomic profiling and fingerprint analysis.

The aim of this study is a comprehensive evaluation of binary similarity measures for the elucidation of patterns among samples of different botanical origin and various metabolomic profiles.

Nine qualitative metabolomic data sets covering a wide range of natural products and metabolomic profiles were applied to assess 44 binary similarity measures for the fingerprinting of plant extracts and natural products. The measures were analyzed by the novel Sum of Ranking Differences method (SRD), searching for the most promising candidates.

SRD and analysis of variance (ANOVA) revealed that Baroni-Urbani-Buser (BUB) and Hawkins-Dotson (HD) similarity coefficients were the most consistent measures while Dice (Di1), Yule (Yu), Russel-Rao (RR), and Consonni-Todeschini 3 (CT3) were ranked the worst. ANOVA revealed that concordantly and intermediately symmetric similarity coefficients are better candidates for metabolomic fingerprinting than the asymmetric and correlation based ones. The fingerprint analysis based on the BUB and HD coefficients and qualitative metabolomic data performed equally well as the quantitative metabolomic profile analysis.

Fingerprint analysis based on the qualitative metabolomic profiles and binary similarity measures proved to be a reliable way in finding patterns in metabolomic data.

Acknowledgement

The project was supported by the National Research, Development and Innovation Office of Hungary under grant numbers K 119269 and KH_17 125608.