Similarity, Diversity - Chemoinformatics

Kurt Varmuza *

Vienna University of Technology (TU Wien), Institute of Statistics and Mathematical Methods in Economics, and Institute of Chemical, Environmental and Bioscience Engineering, A-1040 Vienna, Wiedner Hauptstrasse 7/105-6, Austria

(Received November 5, 2019; Accepted November 21, 2019)

On the occasion of honoring Kimito Funatsu with the 2019 Herman Skolnik Award, aspects of similarity, diversity and complexity are mentioned in relation to chemoinformatics, chemometrics, Japan, and personal encounters with the awardee.

Key Words: spectral database, chemometrics, CHEMICS, structure elucidation

* kurt.varmuza@tuwien.ac.at
Short, Quick, Competent, and Useful

When I visited some years ago the city Kagoshima, where Kimito Funatsu was born and has lived many years, I made a trip to a neighboring peninsula for a close watch of the volcano Sakurajima. It is considered as the most active volcano of Japan, though probably not the most dangerous one. Strolling through a village, about five kilometers from the summit, suddenly a seemingly very huge dark cloud of volcanic ash was rapidly ejected from one of the craters, quickly covering a considerable part of the blue sky. I had wondered to myself - oh, what to do? Obviously, most volcanic activities are difficult to predict, and I was reminded of the similar phrase "like the activity of chemical compounds by QSAR models", having been at a chemoinformatics meeting just a few days prior. First, I looked on local people walking through the street. They completely ignored the volcanic cloud - not even a glance. Second, I remembered an email I had received from Kimito when I had asked him some weeks prior about any reasonable precautions to be taken for a visit of Kagoshima and Sakurajima. His reply was simply "Sakurajima (Kagoshima) is safe." – a response that was typical for him: short, quick, competent, and useful. So, I relaxed and watched the spectacle which disappeared after about half an hour.

Early Chemometrics and CHEMICS

It would have been not difficult to predict the granting of the Herman Skolnik Award to Kimito Funatsu because of his many seminal contributions to chemoinformatics in a period of more than 30 years. He is world-wide one of the core leaders in this field, and has already been awarded from Japanese institutions for his work and successful promotion of computer-aided chemistry in Japanese universities and companies.

The first time this author met Kimito Funatsu was in 1991 at the Toyohashi Institute of Technology. A German-Japanese Workshop on Spectral Databases was held at the location where the CHEMICS system [1] had been developed by Professor Shin-ichi Sasaki and coworkers - with essential contributions by Kimito Funatsu. CHEMICS was and is one of only a few approaches for a systematic and exhaustive chemical structure elucidation based on spectroscopic data and the brutto formula of an unknown organic compound. CHEMICS basically uses NMR and IR spectra for the generation of a so called "goodlist" of substructures that must be present in the unknown, and a "badlist" containing substructures that must be absent. A newly developed structure generator program creates an exhaustive list of all isomers from the given formula - considering the information in the good- and badlist. The resulting molecular candidate structures comprise a complete and systematic solution.

In many practical cases only a small number of candidates remain that can then be checked by additional restrictions known about the query. The origin of this very ambitious approach dates back to the Dendral project in the 1960's which is one of the roots of artificial intelligence. At this time my group worked on the application of multivariate data analysis methods (later called chemometrics) for the recognition of presence/absence of substructures based on low resolution mass spectra. So, I cautiously mentioned my interest in the use of CHEMICS. As things turned out, only a few months later we received from Kimito Funatsu a computer tape (the medium for data storage and exchange in these years) with the CHEMICS software on it. Then, soon afterwards, he personally visited my group at the Vienna University of Technology and introduced to us the use of this sophisticated software. This was impressive and thankworthy. We tried to combine the results of our mass spectra classifiers with the methods implemented in CHEMICS - and realized that the improvement by our mass spectra classification was rather limited [2]. In a later collaboration we found that neural nets applied for variable selection (not for the final classification) improved the performance of mass spectra substructure classifiers considerably [3].

The Japan Chemoinformatics School

Kimito Funatsu organized a highly interesting and successful series of Autumn Schools of Chemoinformatics taking place either at The University of Tokyo or at the Prefectural New Public Hall in Nara. I experienced the honor and pleasure having been invited to give tutorial lectures at some of them. After the event in Tokyo 2015, Kimito joined me on a trip to Kamakura, and once more he offered me many insights to Japanese culture and life. After having a matcha green tea in a quiet tea house we walked up the many stairs of a hill bearing the Great Bell (Engaku-ji temple area) - Kimito easily moving up, myself heavily gasping. We were rewarded by a fantastic clear view to Mount Fuji and discussed various aspects of similarity, diversity and randomness (more elegant: serendipity) – and not only in method development for chemoinformatics. For instance, the highest mountain in Austria (this author's home country) has almost the same height as Mount Fuji - but is not a volcano, serving to remind us that appropriate descriptors are important - not only for chemical structures.
Professor Kimito Funatsu, please accept my admiring congratulations for receiving the famous Herman Skolnik Award. I wish you continuing wise decisions - together with a large piece of luck - for the wide span of similarity, diversity and complexity in your scientific work and for your personal life.

References and Notes

