DUAL DATA DRIVEN SIMCA AS A ONE-CLASS CLASSIFIER

A.L. Pomerantsev[1,2]

[1]Semenov Institute of Chemical Physics RAS, Kosygin 4, Moscow 119991, Russia

[2]Institute of Natural and Technical Systems RAS, Kurortny pr. 99/18, Sochi, 354024, Russia

SIMCA, the method of soft independent modeling of class analogy, was proposed about 40 years ago [1]. This approach is a natural extension of a well-known method of principal component analysis (PCA). The main SIMCA function is qualitative data analysis that can be formulated as following. Let us consider a class of objects **X** (substance, materials, products). A set of analytical measurements (spectra, chromatograms, etc.) represents the class of objects accounting for a possible variability. Suppose we have a new object **y**, for which we have to decide whether **y** belongs to class **X,** or not. An example of such a problem is the recognition of counterfeit drugs [2]. SIMCA has been revised repeatedly [3-6]. Today this method is very popular in analytical chemistry (chemometrics), but almost unknown outside. SIMCA provides a unique opportunity to make classification accounting both for the Type-I error α (false rejection) and the Type-II error β (false acceptance), however this is used extremely rare. The SIMCA theoretical base is thoroughly developed, but most of the analytical studies contain gross errors, which are repeated from publication to publication.

The presentation is going to bridge the gap and to provide a general SIMCA concept as a data driven method. The following items will be considered

- How PCA relates to SIMCA
- The distance measures in use: score distance, orthogonal distance, total distance
- What statistics are used in SIMCA and how this statistics are distributed
- How to make the decision at a given the Type I error α
- How to calculate the Type II error β

Presentation is illustrated with simple examples.

References

1. Wold S. *Pattern Recognition* 1976; **8**: 127-139.
2. Rodionova OY, Pomerantsev AL, *Trends Anal. Chem*., 2010; **29**: 781-938.
3. Nomikos P, MacGregor JF. *Technometrics* 1995; **37**: 41–59
4. Pomerantsev AL, *J. Chemometrics* 2008; **22**: 601-609
5. Hubert M, Rousseeuw PJ, Vanden Branden K. *Technometrics*, 2005; **47**: 64-79
6. Pomerantsev AL, Rodionova OY, *J. Chemometrics* 2013 (DOI: 10.1002/cem.2506)