Trends in Analytical Chemistry 69 (2015) 34–51

Contents lists available at ScienceDirect

Trends in Analytical Chemistry

journal homepage: www.elsevier.com/locate/trac

Chemometrics and qualitative analysis have a vibrant relationship

Ewa Szymańska [a,b,*], Jan Gerretzen [a,b], Jasper Engel [b], Brigitte Geurts [b], Lionel Blanchet [b,c], Lutgarde M.C. Buydens [b]

[a] TI-COAST, Science Park 904, 1098 X.H. Amsterdam, The Netherlands
[b] Institute for Molecules and Materials (IMM), Radboud University, P.O. Box 9010, 6500 GL Nijmegen, The Netherlands
[c] Department of Biochemistry, Radboud Institute for Molecular Life Sciences, Radboud University Medical Center, P.O. Box 9101, NL-6500 HB Nijmegen, The Netherlands

A R T I C L E I N F O

Keywords:
Biomarker identification
Chemometrics
Compound identification
High-dimensional data
Qualitative analysis
Qualitative data
Qualitative goal
Quantitative data
Sample classification
Validation

A B S T R A C T

In analytical chemistry, qualitative analysis is often associated with compound identification, while chemometrics offers a wide spectrum of data-analysis methods that extend the application of qualitative analysis beyond compound identification. All chemical analyses that have a qualitative goal can or should be considered as qualitative chemical analysis. Thanks to chemometrics, both qualitative and quantitative data can be included in qualitative analysis and modeled towards a qualitative analysis goal. We provide an extensive overview on the vibrant relationship between chemometrics and qualitative analysis. It includes a description of chemometric methods, their real-life applications in qualitative analysis, challenges and possible solutions. Undoubtedly, the role of chemometrics will become pivotal in the future when more possibilities of qualitative analysis will be explored and new chemometric approaches will be developed for high-dimensional data.

Contents

Abbreviations: ANN, Artificial Neuronal Network; CART, Classification and Regression Trees; CVA, Canonical Variate Analysis; GA, Genetic Algorithms; HCA, Hierarchical Cluster Analysis; ICA, Independent Component Analysis; k-NN, k-Nearest Neighbors; LDA, Linear Discriminant Analysis; MCR, Multivariate Curve Resolution; MDS, Multidimensional Scaling; MSPC, Multivariate Statistical Process Control; PARAFAC, Parallel Factor Analysis; PCA, Principal Component Analysis; PF, Potential Functions; PLS-DA, Partial Least Squares-Discriminant Analysis; PP, Projection Pursuit; QDA, Quadratic Discriminant Analysis; RF, Random Forest; SIMCA, Soft-Independent Modeling of Class Analogy; SOMs, Self-Organizing Maps; SVM, Support Vector Machines; UNEQ, Unequal Class-Modeling.

* Corresponding author. Tel.: +31 24 3653180; fax: +31 24 3652653.
E-mail address: chemometrics@science.ru.nl (E. Szymańska).

http://dx.doi.org/10.1016/j.trac.2015.02.015
0165-9936/

1. Introduction: scope of qualitative analysis

According to the definition of the International Union of Pure and Applied Chemistry (IUPAC [1],) *qualitative analysis* refers to analyses in which substances are identified or classified on the basis of their chemical or physical properties, such as chemical reactivity, solubility, molecular weight, melting point, radiative properties (emission, absorption), mass spectra, and nuclear half-life. This definition originates from 1995. Nowadays, qualitative analysis applies to not only identification of single chemical compounds but also many other qualitative applications of analytical chemistry. One can think about the following examples:

a. urine samples coming from healthy people and cancer patients can be qualified based on metabolic profiles measured by capillary electrophoresis [2];
b. rice samples from the European Union can be identified among rice samples from other locations based on NMR spectra [3];
c. biomarkers of a cholesterol-lowering food intervention can be identified among blood-lipid metabolites measured by LC-MS [4];
d. a set of wavelengths discriminating one type of beer from another can be selected from NIR spectra of different beer samples [5]; and,
e. presence or absence of toxic compounds in urine samples can be determined by studying NMR spectra [6].

In each of these examples, analytical measurements are employed to qualify samples, compounds, or groups of compounds. In the recent book of Milman [7], identification/classification/authentication of, e.g., foodstuffs, products, specimens, and materials, is referred to as qualitative analysis II. An extension of the IUPAC definition is therefore required to cover different aspects and applications of qualitative analysis. For the purpose of this review, we define such an extension based on the framework of chemical analysis presented in the next paragraph.

We position qualitative analysis in the general framework of chemical analysis as shown in Fig. 1. The goal of chemical analysis can be either qualitative or quantitative (step s1 in Fig. 1). A qualitative goal means that the objects of interest (e.g., samples, compounds) will be qualified in one way or another; e.g., urine samples are qualified as coming from either healthy or diseased based on their metabolic profile (see example a). However, a quantitative goal focuses on the quantity/quantities of compounds (e.g., the concentration of alanine in a plasma sample). After defining the goal, samples are collected and analyzed with an appropriate analytical method (steps s2–s3 in Fig. 1). The outcome of such analysis is analytical chemical data (e.g., a chromatogram or an NMR spectrum) (step s4). Depending on the measurement scale, data can be either quantitative or qualitative.

In some cases, the same data can be treated as both quantitative and qualitative (e.g., chromatographic data can be treated as qualitative when only the presence or absence of a peak at a specified retention time is of interest, or as quantitative when the intensities of specified peaks are of interest). After the data have been collected, exploration and translation of the analytical chemical data into the required qualitative information takes place. This is an essential step where chemometric methods play a pivotal role (step s5 in Fig. 1). Since the acquired data are often multivariate in nature, the selection and the application of an appropriate chemometric method is an essential step and is not always straightforward. Two determining factors are the goal of the analysis and the type of data (i.e., quantitative or qualitative).

For the purposes of this review, the definition and the scope of qualitative analysis are specified as the analysis of both qualitative and/or quantitative chemical data towards a qualitative goal (red box in Fig. 1). It should be stressed that this definition is broader than the definition of IUPAC because a) a qualitative goal is not limited to single-compound identification and b) quantitative data are also included in this definition of qualitative analysis.

For many analytical chemists, analysis of qualitative data towards a quantitative goal is also an example of a qualitative analysis. One typical example is quantitative structure activity relationship (QSAR) models [8], where chemical structures are used to quantify the activity of a given compound. However, in this review, we explicitly focus on analysis towards a qualitative goal so we do not consider QSAR modeling.

In this review, we describe the vibrant relationship between chemometrics and qualitative analysis. We give an overview of chemometric methods for qualitative analysis, their real-life applications, challenges and possible solutions. First, in Section 2, the main differences between qualitative and quantitative data used in qualitative analysis are explained. Next, we discuss chemometric approaches appropriate for the analysis of qualitative data (Section 3). Third, we detail chemometric approaches suitable for the analysis of quantitative data – towards a qualitative goal – in Section