Food Control 86 (2018) 283–293

Contents lists available at ScienceDirect

Food Control

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

Review

An overview of multivariate qualitative methods for food fraud detection

M. Pilar Callao*, Itziar Ruisánchez

Chemometrics, Qualimetric and Nanosensors Group, Department of Analytical and Organic Chemistry, Rovira i Virgili University, Marcel·lí Domingo s/n, 43007 Tarragona, Spain

ARTICLE INFO

Article history:
Received 21 September 2017
Received in revised form
20 November 2017
Accepted 23 November 2017

Keywords:
Multivariate qualitative method
Food authentication
Food adulteration
Classification
Data fusion
Validation

ABSTRACT

Multivariate qualitative methods are an analytical strategy for addressing problems related to food fraud that cannot be solved with just one variable. Some examples are sample authentication since the required response is complex in nature and sample adulteration, when knowing the concentration of adulterant is not looked for. Establishing a multivariate qualitative method involves several steps: data collection, pre-treatment, exploration techniques, classification techniques, and method validation. When more than one data source is available, data fusion can be apply to improve the results of a single technique.

This review describes the state of the art of multivariate qualitative analysis for determining food fraud, and differentiates between authentication and adulteration. All the mentioned steps are discussed and, as example, recently published papers are commented.

Contents

1. Introduction

Qualitative methods are by no means new. Although they are not used in routine laboratory tasks as much as quantitative methods, they are currently on the rise and have been attracting increasingly greater interest, mainly for their screening potential.

Qualitative methods can be classified using several criteria but in all cases they are used in problems that require a binary response (yes/no). If response was achieved from multiple non-specific signals, a multivariate classification approach is required. These strategy is also referred as non-target analysis since the data set is used as a fingerprint of the sample.

According to the literature, multivariate qualitative methods are increasingly used in many fields (chemistry, process monitoring, etc.). Of course, multivariate classification is becoming increasingly important in food science too (Ballabio & Todeschini, 2009, chap. 4). In this paper, we focus more precisely on multivariate qualitative methods for problems of food fraud. In food fraud analysis, there are two main problems: a) authenticating the origin of a product in terms of geographical or botanical/animal provenance, or the manufacturing process, b) proving the absence of adulteration or the addition of a non-declared substance.

* Corresponding author.
E-mail address: mariapilar.callao@urv.cat (M.P. Callao).

https://doi.org/10.1016/j.foodcont.2017.11.034
0956-7135/

As far as product authentication is concerned, in many countries there are laws that require agricultural products to have information about their geographical origin on the labels. The EU has encouraged the use of labelling to identify products by introducing regulations, first in 1992 and more recently in 2006 (EU regulations 510/2006, 509/2009 and 1898/2006). Those regulations define the following geographical indications for food products: protected designation of origin (PDO), protected geographical indication (PGI) and traditional specialities guaranteed (TSG). The use of geographical indications implies market recognition and it is related to the price of the product. To solve the problem of authentication, the response required is qualitative; that is, binary (yes/no; belongs/does not belong, etc.). However, a single signal often cannot solve the problem, so a multivariate approach is usually required.

The second problem, food adulteration, is attracting increasing attention because it is an emerging risk, given the complex and global nature of food supply chains. One of the major concerns about adulteration is that it may involve a health risk or economic benefit. Food adulteration problems can be solved in two ways: if the adulterant is known, a quantitative analysis is usually carried out but, if it is not, a qualitative analysis (it is or it is not adulterated) may be satisfactory.

A bibliographic search of the last five years shows how keywords such as “food authentication” or “food adulteration” and “classification” were increasingly found in scientific articles. They mainly refer to the use of classification techniques with a multivariate signal provided by different instrumental techniques. Recently, several reviews have been published on specific instrumental techniques that are used with a chemometric approach for food analysis (Bosque-Sendra, Cuadros-Rodriguez, Ruiz-Samblas, & De la Mata, 2012; Casale & Simonetti, 2014; Danezis, Tsagkaris, Camin, Brusic, & Georgiou, 2016; Domingo, Tirelli, Nunes, Guerreiro, & Pinto, 2014), the use of chemometric techniques for specific food analysis (Camiña, Pellerano, & Marchevsky, 2012; Domingo et al., 2014; Esslinger, Riedl, & Fauhl-Hassek, 2014; Haddi et al., 2014; Kamal & Karoui, 2015; Nascimento, Santos, Pereira-Filho, & Rocha, 2017), or the metabolomic analysis of food (Cubero-Leon, Peñalver, & Maquet, 2014).

This overview focused on the development of multivariate qualitative methods for the detection of food fraud. Fig. 1 schematically presents an overall protocol for this purpose. It should be noted that the analytical determinations that give rise to the data set are mainly instrumental measures that provide multiple data for each sample analysed (i.e. absorbance at different wavelengths), although they can also be independent measures from different techniques (i.e. pH, conductivity, etc.). The former are more common, because the experimental cost is very small.

The paper has been divided into sections that correspond to the different steps implemented in a multivariate qualitative method. Section 2 (exploratory analysis) and section 3 (classification techniques) are the steps that have been studied most, so the main characteristics of the different approaches will be commented. Section 4 (data fusion) is the step more recently introduced in multivariate qualitative analysis. Section 5 focuses on the validation step. Recently some studies (Lopez, M.I. et al., 2015; Riedl, J. et al., 2015) deals with it, although further research is required to develop unified protocols. In each section, the chemometric techniques are briefly described, although for more in-depth explanations the reader is addressed to the basic bibliography.

2. Exploratory analysis

The exploratory or unsupervised analysis provide information about the relationship between samples, between variables and/or between samples and variables. Various tools can be used and their theoretical basis has been well explained in many scientific articles and recent books on chemometrics (Esbensen & Geladi, 2009, chap. 2.13; Li Vigni, Durante, & Cocchi, 2013).

Information about the relationship between samples reveals whether there are natural groups or trends in sample distribution that are consistent with prior knowledge about them. For example, if a strategy is established for detecting authentication and both authentic and non authentic samples are submitted to an unsupervised analysis, they should present a distribution that shows some tendencies. If there are not tendencies, the characterization of the samples must be not adequate and the experimentation carried out must be redefined. In addition, unsupervised techniques make it possible to detect the presence of possible outliers: i.e. samples distributed differently and separate from the main group. These samples should be rejected as they can have a negative impact on the use of supervised techniques.

The relationship between variables shows which of them give complementary information and which give similar or redundant information. On the other hand the relationship between samples and variables indicates which variables are important (and which are not) for distinguishing groups of samples. This type of information can be valuable to simplify the database or, in some cases, to reduce experimentation.

The most popular unsupervised exploratory technique is based on the well-known principal components analysis (PCA) (Esbensen & Geladi, 2009, chap. 2.13; Li Vigni et al., 2013). PCA generates new variables as a linear combination of the original variables. These new variables retain maximum information from the original data matrix and are called principal components (PCs). The first PC is the one that retains most explained variance (more data information) while the second PC explains the information that is not modelled by the first PCs, and so on. When it is used as exploratory technique, the information from the two or three first PC's are plot. So, sample and variable distribution are showed. Its main limitation is when the first PC's do not contain enough information.

Other exploratory techniques are cluster analysis (CA) (Lee & Yang, 2009, chap. 2.17), in which samples (or variables) are linked to others according to their similarity. Groups considering similarity values are defined. The main limitation of this technique is that it does not show the overall relationship between all the samples but only between the ones that are close together. Neither does it give any information about the relationship between samples and variables. On the other hand, it uses all the information contained in the data and can be considered to complement the PCA representation.

As Table 1 shows, most authentication or adulteration studies use the PCA technique before applying a classification technique. Some studies also use cluster analysis techniques (Mir-Marqués, Elvira-Sáez, Cervera, Garrigues, & De la Guardia, 2016; Azevedo, M.S. et al., 2017).

Some of the studies reviewed only present a PCA exploratory analysis, and interpret both the scores and the loading plot (Boggia, Casolino, Hysenaj, Oliveri, & Zunin, 2013; Dahimi et al., 2014; Malheiro, Pinho, Soares, Ferreira, & Baptista, 2013; Üçüncüoğlu, İlaslan, Boyacı, & Özay, 2013). For instance, PCA was used in the study of six fresh wild mushroom species for taxonomical and authentication purposes (Malheiro et al., 2013). The authors used the loading plot to identify the volatile secondary metabolites (11 volatile compounds out of forty-six) that characterize each mushroom species and which have highest power of discrimination. These compounds seem to play a crucial biomarker role in the characterization of the six wild species of mushrooms.

Similarly, a screening method was proposed to detect pomegranate juice adulteration by the addition of cheaper fruit juices