



^1H NMR and UV-visible data fusion for determining Sudan dyes in culinary spices

Carolina V. Di Anibal, M. Pilar Callao, Itziar Ruisánchez*

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

ARTICLE INFO

Article history:

Received 15 November 2010
Received in revised form 1 February 2011
Accepted 14 February 2011
Available online 24 February 2011

Keywords:

Variable level data fusion
Decision level data fusion
UV-visible
 ^1H NMR
Fuzzy aggregation connectives
Food adulteration

ABSTRACT

Two data fusion strategies (variable and decision level) combined with a multivariate classification approach (Partial Least Squares-Discriminant Analysis, PLS-DA) have been applied to get benefits from the synergistic effect of the information obtained from two spectroscopic techniques: UV-visible and ^1H NMR. Variable level data fusion consists of merging the spectra obtained from each spectroscopic technique in what is called “meta-spectrum” and then applying the classification technique. Decision level data fusion combines the results of individually applying the classification technique in each spectroscopic technique. Among the possible ways of combinations, we have used the fuzzy aggregation connective operators. This procedure has been applied to determine banned dyes (Sudan III and IV) in culinary spices. The results show that data fusion is an effective strategy since the classification results are better than the individual ones: between 80 and 100% for the individual techniques and between 97 and 100% with the two fusion strategies.

© 2011 Elsevier B.V. All rights reserved.

1. Introduction

Nowadays, food quality and safety characterization is still one of the key issues whenever dealing with foodstuffs and great effort has been devoted to the detection of hazardous additives. The application of spectroscopic techniques has become a usual tool in food analysis [1] and requires the use and development of chemometrics tools in order to display and interpret vast amounts of data. There are some approaches that couple or merge data from two or more analytical techniques to improve multivariate data interpretation. This procedure goes under different names: data fusion, analysis of coupled or linked data, multiset or multiblock data analysis and integrative data analysis [2], among others. In this manuscript, we shall use the term data fusion as it is the term most commonly used by data analysts.

Data fusion has been applied in a variety of fields: for example, the combination of electronic noses and spectroscopic techniques to authenticate olive oil [3] and white grape must [4], and to determine sensory attributes in red wines [5]. Spectroscopic techniques have also been fused to identify pigments in works of art [6,7] and cultivars of extra virgin olive oils [8]. Most fused data comes from Infrared (NIR, FT-IR, MIR), UV-visible, Raman, Fluorescence, and Mass Spectrometry. One of the emerging research areas in which

data fusion is applied is known as “metabolomics” or “metabonomics”, the goal of which is to obtain information from such highly complex samples as biofluids, cells and tissues [9,10].

Previous studies have shown that the UV-visible [11] and High-Resolution ^1H Nuclear Magnetic Resonance (^1H NMR) [12] spectroscopic techniques coupled with multivariate classification techniques are well suited for determining the possible adulteration of commercial spices with Sudan dyes (I–IV). Based on those results, the main objective of the present study is to evaluate the combination of both techniques to improve the classification results. In addition, as different concentrations of Sudan dyes can be found in adulterated spices [13], in this paper we will explore the classification ability when the samples have lower concentration levels than those studied in the papers mentioned above. Also, as most of the classification errors were obtained with samples adulterated with Sudan III and IV, the present study focuses on these two dyes. The idea is to get benefits of the possible synergism that two techniques as UV-visible and NMR could have each other; due to the fact that both are based in different fundamentals and give different analytical signals, which allows thinking that the information provided by each one could be complementary.

The data provided by the two spectroscopic techniques have been processed separately and jointly by two data fusion strategies: variable and decision level data fusion. So, the overall performance of the classification process is evaluated through the well known classification technique Partial Least Squares-Discriminant Analysis (PLS-DA) for each individual spectroscopic technique and the fusion process.

* Corresponding author at: Department of Analytical and Organic Chemistry, Rovira i Virgili University, Campus Sescelades, Marcel·lí Domingo s/n, 43007 Tarragona, Spain. Tel.: +34 977558490; fax: +34 977558446.

E-mail address: itziar.ruisanchez@urv.cat (I. Ruisánchez).

2. Materials and methods

2.1. Samples

A total of 35 spices from different common markets were studied. For UV-visible analysis, each spice was extracted with acetonitrile and the obtained extract was twice filtered. For NMR, samples were dissolved in deuterated chloroform and once filtered. Samples contaminated with Sudan III and IV were prepared by spiking the non-contaminated samples at three concentration levels: 1.4, 3.6 and 7.1 g/kg. In the end, then, three classes were defined: class 1 contained the 35 non-adulterated spices, class 2 contained a total of 105 samples adulterated with Sudan III corresponding to the three concentration levels (35 samples each one) and class 3 contained 105 samples corresponding to the three concentration levels adulterated with Sudan IV. More details about the sample treatment and experimental section can be found in previous studies [11,12].

2.2. Spectrometric techniques and dataset

¹H NMR spectra were acquired at 600.13 MHz on a Bruker Avance III-600 spectrometer, equipped with an inverse TCI 5 mm cryoprobe[®]. One dimensional pulse experiments were carried out using a 90° pulse sequence (zg). For each sample, eight scans of 9.6 kHz of spectral width were collected at 300K into 64K data points. A recycling delay time of 15 s was applied between scans to ensure fully relaxation. Exponential line broadening of 0.3 Hz was applied before Fourier transformation and the NMR spectra acquired were phased, baseline-corrected (5th order polynomial adjustment) and calibrated by setting the CDCl₃ peak at 7.27 ppm (TopSpin 2.1, Bruker Biospin, Rheinstetten, Germany). UV-visible measurements were made by an Agilent 8453 UV-visible spectrophotometer (Agilent Technologies Inc., Palo Alto, CA, USA) equipped with a diode array detector (DAD) and ChemStation Software (ChemStation Rev. A. 08.03). Each sample was measured against solvent as a blank in a 1-cm pathlength quartz cell and with a spectral resolution of 1 nm.

The UV-visible spectra were acquired between 260 and 600 nm and had a total of 341 variables. The spectral region for NMR is located between 0.5 and 8.9 ppm and a range corresponding to the solvent signal (centred at 7.26 ppm) was removed, so finally, there were a total of 5698 variables.

The dataset (245 samples) is divided into a training and test set. The test set was generated by leaving out a 14% of the samples from class 1 and from each of the three concentration levels included in classes 2 and 3. The selection criterion is based on the PCA scores plot distribution from both UV-visible and NMR data. Finally, the training set consisted of 210 samples and the test set of 35.

3. Chemometrics tools

3.1. Software

All chemometrics treatment was made with Matlab 6.5 Software (The MathWorks, Natick, MA) and PLS.Toolbox 3.5 (Eigenvector Research Incorporated).

3.2. Partial Least Squares-Discriminant Analysis

PLS-DA is the classical PLS regression technique adapted to a supervised classification task. A regression model is calculated that relates the independent variables (e.g. spectra) to an integer “y” that designates the class of the sample, with a binary response encoded as {1 0 0} meaning that a sample belongs to class 1; {0 1 0} to class 2 and {0 0 1} to class 3. The model predicts the class for each sample

based on a value from zero to one. A value close to zero indicates that the sample is not in the modelled class, while a value closer to one indicates that it is. A threshold between 0 and 1 (above which a sample is considered part of the class) is calculated using Bayesian statistics [14]. The Bayesian threshold assumes that the “y” PLS predicted values are normally distributed and the threshold is selected at the y value at which the number of false positives and false negatives is minimized. More details of the PLS-DA technique can be found in the literature [15,16].

The optimal number of latent variables (LVs) to be included in each model was chosen using leave-one-out cross-validation to minimize the root mean square-cross validation error (RMSECV) for each class. At the end, this number is selected through a compromise between the optimal value for each class.

3.3. Data fusion

Two levels of data fusion architectures are investigated in this paper: variable and decision level data fusion.

3.3.1. Variable level data fusion

Variable level fusion concatenates the variables into a single vector, which is called a “meta-spectrum”. Data must be balanced (all variables in the same scale) prior to the fusion process, so in our particular case only the NMR variables are normalized since the UV-visible intensity values are already between 0 and 1. If the number of concatenated variables is quite high, a variable selection is required. Of the various selection approaches, interval Partial Least Squares (iPLS) was used here [17]. We are not going to describe the iPLS methodology in detail, merely point out that it investigates the influential zones of the spectra that contain the most discriminating predictors, and calculates local PLS-DA models in pre-fixed narrow intervals.

3.3.2. Decision level data fusion

Decision level data fusion combines the classification results obtained from each individual technique. In this study, the PLS-DA classification results are fused using the fuzzy set theory which implements fuzzy aggregation connective operators. Fuzzy theory, introduced by Zadeh [18], is a powerful and general technology for processing information.

A fuzzy set allows membership values between 0 and 1, so in our case the PLS-DA class assignment values are normalized to the interval [0,1] through a simple rescaling such as the following Eq. [19]:

$$m_{ij} = \frac{x_{ij} - \min(x_{ij})}{\max(x_{ij}) - \min(x_{ij})}$$

where (m_{ij}) is the normalized class assignment value, x_{ij} is the PLS-DA class assignment value for the i th sample in the j th class and min and max are the minimum and maximum class assignment values in each j th class considering all samples, respectively.

Of the wide range of fuzzy connectives and aggregation operators that are available, we chose four aggregation connective operators which belong to the class of context independent constant behavior (CICB) operators [20]. The aggregation operators are: Minimum, Maximum, Product and Arithmetic Average. Considering the PLS-DA normalized assignment values obtained with each technique, the minimum and maximum are identified and the product and average are calculated. For the sake of clarity, two examples will be shown in the results section. To obtain the “ensemble decision” of each operator, the maximum value of the three possible classes is chosen [21]. The sample is finally assigned by the majority vote provided by all the fuzzy operators in the “ensemble decision”.

Download English Version:

<https://daneshyari.com/en/article/1244177>

Download Persian Version:

<https://daneshyari.com/article/1244177>

[Daneshyari.com](https://daneshyari.com)