Sensors and Actuators B 191 (2014) 67–74

Contents lists available at ScienceDirect

Sensors and Actuators B: Chemical

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

On the application of simple matrix methods for electronic tongue data processing: Case study with black tea samples

Irina Yaroshenko[a], Dmitry Kirsanov[a,*], Liudmila Kartsova[a], Nabarun Bhattacharyya[b], Subrata Sarkar[b], Andrey Legin[a]

[a] Chemistry Department, St. Petersburg State University, Universitetskaya nab. 7/9, Mendeleev Center, 199034 St. Petersburg, Russia
[b] Centre for Development of Advanced Computing (C-DAC), Kolkata, India

ARTICLE INFO

Article history:
Received 22 March 2013
Received in revised form 3 September 2013
Accepted 23 September 2013
Available online 1 October 2013

Keywords:
Electronic tongue
Data analysis
Matrix correlation
Black tea

ABSTRACT

The paper is devoted to the application of several simple mathematical methods that can be used for analysis of multisensor system data. RV and modified RV′ matrix correlation coefficients, Tucker's congruency coefficients and canonical correlation analysis can be effectively employed to reveal the similarity degree shared between two different data sets obtained for the same samples. This approach can be particularly useful when the number of available samples is too small to construct and validate quantitative regression models, but still the researcher is interested in getting some numerical estimate of the data quality to judge on electronic tongue applicability for particular analytical tasks. As a case study for this work we analyzed the data from potentiometric and voltammetric electronic tongues acquired during the analysis of black tea samples. These samples were also assessed with standard physicochemical methods and professional sensory panel. The relationships between these four data sets were studied with various matrix correlation techniques.

1. Introduction

Electronic tongue methodology is getting more and more popular in analytical chemistry for food and beverage analysis [1,2], pharmaceutical analysis [3,4], water quality control [5,6]. The main idea of this approach is an employment of an array of chemical sensors with high cross-sensitivity in a complex liquid media and subsequent processing of the obtained analytical signal by means of multivariate data processing techniques [7]. The sensors for such arrays can be based on various principles and up to date there are numerous reports on application of potentiometric [8], voltammetric [9], impedimetric [10], optical [11], surface acoustic wave [12] and other measuring methods. The cross-sensitivity of the sensors assumes that each one of them can yield a response to the number of substances in the sample and this leads to unresolved analytical signal of such array. Multivariate data processing techniques are required to produce qualitative and quantitative chemical information from these signals. In certain cases the proper choice of data processing method can be of the same importance as the choice of the sensing platform itself. While the development of novel sensors and sensing materials received a lot of attention through the last decade, the second indispensable part of ET systems – multivariate data processing is usually handled with leftover principle. This leads to the fact the repertoire of chemometric techniques in use by ET community is quite limited. The vast majority of the papers devoted to the electronic tongues are dealing with PCA (Principal Component Analysis), PLS (Projection on Latent Structures) and various kinds of ANN (Artificial Neural Networks) as the main data processing techniques. This observation is also confirmed with extensive literature survey in [13]. Although these methods proved to be very powerful and reliable for ET data processing the use of the other multivariate approaches and algorithms should not be neglected.

Typical situation in ET field is that the researcher has two sets of data: the first one is the data from the instrument where i samples are analyzed by j sensors or variables to be more general ($i \times j$ matrix) and the second one is a result of some reference analysis of the same i samples. The latter can be the k concentrations of certain substances obtained from chromatographic or other methods or in many cases the scores from human sensory panel with the number of k taste descriptors assessed ($i \times k$ matrix). The researcher is usually interested to find out if there are some relations between these two data sets and, as an ultimate goal, if it is possible to predict the reference values (concentrations, sensory panel scores, etc.) from the set of j independent variables (the signals of ET sensors in the studied samples). It is not infrequent that the researcher has only a limited number of samples available for a study, this could be due to the specific requirements of the experiment (e.g. only some anomalous samples are interesting and this anomaly is quite rare), the price of the reference analysis is too high to manage large

* Corresponding author. Tel.: +7 921 333 12 46.
E-mail address: d.kirsanov@gmail.com (D. Kirsanov).

0925-4005/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.snb.2013.09.093

sample set, etc. In this case a construction of a reliable regression models for prediction of a particular parameters of interest cannot be feasible. This is especially true for validation procedure; the only available option with a small sample set could be a cross-validation which is widely criticized for over-optimistic results [14–16]. However, still the researcher wants to have some numerical estimation of the ET data quality and ET applicability for the chosen analytical task. In such cases the mathematical instruments popular, e.g. in psychometrics can be adopted.

There are several ways to study relations between data sets (matrices) and probably the simplest one is RV matrix correlation coefficient which is a generalization of variable correlation for multivariate case [17]. The values of RV coefficient may vary from zero (no correlation) to one (perfect correlation). Due to the possibility of obtaining relatively high RV values in certain cases when two data sets are not in fact correlated (chance correlation) the modified RV coefficients were recently introduced [18]. There is also a well-established procedure to study the similarity of latent variables extracted e.g. with PCA for two data sets, the value of Tucker's congruence coefficients [19,20] can be used for that. A canonical correlation analysis (CCA) [21] is also an effective tool for this purpose. The advantage of the CCA is that its results can be easily visualized in the form of similarity maps displaying the same objects (samples) in the domains of ET and referent measurements. There are not too many reports on the use of these methods for data processing in chemistry [22–24], however they can be very useful in spite of their simplicity.

The purpose of this research is to introduce the above mentioned methods for sensor community and to show the advantages of applying certain simple mathematical methods for ET data processing. Most of these methods involve very basic matrix algebra and the results can be easily computed in some common data manipulation software, like e.g. MS Excel or freely available *R* [25]. Nevertheless such results can provide many useful insights into the data.

As a case study for this research we use several sets of black tea samples studied by potentiometric and voltammetric ETs as well as by physicochemical methods and professional sensory panel. There are several reports on tea analysis with help of various ETs [26–30] however black tea is still very interesting research object being one of the most consumed beverages in the world.

2. Materials and methods

2.1. Black tea samples preparation

Ten black tea samples were collected for this study. The samples were prepared by CTC technique. CTC (Crush, Tear, Curl) is a method of black tea processing which involves crushing, tearing and curling of tea leaves (hence the name). This procedure yields very strongly flavored quickly infusing teas that can be boiled or used in instant teabags. Nowadays this method is the most popular in India (over 80% of tea production is of the CTC type). The samples for the study were provided by the laboratory of Tocklai Experimental Station (Assam, India), one of the oldest tea research centers all over the world. All samples are originated from Assam tea gardens and belong to the same quality rank.

Preparation of tea samples for analysis by potentiometric multisensor system was done in the following way: 2 g of dry tea sample were brewed in 100 mL of freshly boiled distilled water for 5 min, infusion was allowed to cool down for 20 min to room temperature (25 °C). After that 30 ml of broth were diluted with 70 ml of distilled water, the resulting solution was employed for potentiometric measurements. Only liquid part of sample was used for measurements and all leaves particles were not transferred into the measurement cell.

For voltammetric multisensor analysis 750 mg of dry tea sample were brewed for 5 min in 150 mL of distilled water. The tea leaves were separated on a paper filter and measurements were performed after infusion was allowed to cool down to room temperature (25 °C) for 20 min.

The differences in sample preparation procedures for ET measurements are due to the fact that both research groups form India (voltammetric ET) and from Russia (potentiometric ET) have their own established protocols for dealing with tea samples and these protocols were optimized during the preliminary experiments. Fig. 1 shows typical response curves from two ETs in black tea samples.

2.2. Potentiometric and voltammetric multisensor systems

The potentiometric multisensor system was comprised of 19 cross-sensitive chemical sensors, 10 anion-sensitive with polymeric membranes, 5 cation-sensitive with polymeric membranes and 4 chalcogenide glass sensors with RedOx sensitivity. All sensors were purchased from Sensor Systems, LLC (St. Petersburg, Russia). Besides that the system contained standard pH glass electrode and Ag/AgCl reference electrode, both from ZIP (Gomel, Belorussia). All sensors were connected with shielded wires to the 32-channel high input impedance digital mV-meter. Potential readings were recorded with 0.1 mV precision in a custom made software installed on a Windows PC. Measurement time in each tea sample was 3 min, and the averaged values of the sensor response during last 20 sec were employed for further data processing. After that the sensors were washed with 3 portions of distilled water 3 min each. All samples were measured three times and these results were averaged over the replicas for further processing, the resulted data matrix from potentiometric multisensor system was 10 samples × 20 sensors (19 cross-sensitive + 1 pH sensor). Due to the nature of the sensors the resulted data contained information on ionic species in the samples, both inorganic and organic and also on RedOx pairs.

The voltammetric multisensor system was an array of six working electrodes, made of gold, copper, palladium, iron, nickel and glassy carbon; a platinum counter electrode (PH Ionics, India); and Ag/AgCl reference electrode (Gamry Instruments Inc., USA). This experimental setup was designed in Centre for Development of Advanced Computing (CDAC, Kolkata, India) and was already described in literature [31,32]. The voltammetric scans were performed with custom made potentiostat based on the data acquisition card USB 6008 (National Instruments, USA) in the voltage range ± 1.5 V. The scan rate was 0.1 V/s with a step potential 9 mV, unfolded cyclic voltammograms were employed for data processing.

Before each measurement session the electrodes were kept in one of the samples for conditioning purposes. The electrodes were rinsed with distilled water after each sample. The values from each electrode were recorded in data files. In this case the data matrix from voltammetric measurements was 10 samples × 1092 current readings (182 values for each working electrode). These data contained information on all substances which can be reduced/oxidized under described experimental conditions.

2.3. Reference data on black tea samples

The data from reference analytical method for the tea samples under study were provided by the laboratory of Tocklai Experimental Station (Assam, India). These were the results from physicochemical methods: contents of theaflavins (TF) and thearubigins (TR), total color and brightness estimated by colorimetric method with Cary 50 UV-VIS Spectrophotometer (Varian, USA) as