Available online at www.sciencedirect.com
SCIENCE DIRECT®
Talanta 67 (2005) 590–596
Talanta
www.elsevier.com/locate/talanta

Direct and two-stage data analysis procedures based on PCA, PLS-DA and ANN for ISE-based electronic tongue—Effect of supervised feature extraction

P. Ciosek [a,*], Z. Brzózka [a], W. Wróblewski [a], E. Martinelli [b], C. Di Natale [b], A. D'Amico [b]

[a] *Department of Analytical Chemistry, Warsaw University of Technology, Noakowskiego 3, 00-664 Warsaw, Poland*
[b] *Department of Electronic Engineering, University of Rome "Tor Vergata" Via di Tor Vergata, 00133 Rome, Italy*

Received 13 August 2004; received in revised form 15 February 2005; accepted 14 March 2005
Available online 14 April 2005

Abstract

A novel strategy of data analysis for artificial taste and odour systems is presented in this work. It is demonstrated that using a supervised method also in feature extraction phase enhances fruit juice classification capability of sensor array developed at Warsaw University of Technology. Comparison of direct processing (raw data processed by Artificial Neural Network (ANN), raw data processed by Partial Least Squares-Discriminant Analysis (PLS-DA)) and two-stage processing (Principal Components Analysis (PCA) outputs processed by ANN, PLS-DA outputs processed by ANN) is presented. It is shown that considerable increase of classification capability occurred in the case of the new method proposed by the authors.

Keywords: Electronic tongue; Sensor array; Ion-selective electrodes (ISEs); Pattern recognition

1. Introduction

Since the first work dealing with electronic nose [1], many odour-sensing systems have been presented [2]. In 1985, the first system for liquid analysis was described by Otto and Thomas [3]. During last 20 years, only a few devices for liquids called "electronic tongues" have been presented. The majority of those systems are based on electrochemistry [4–6]; however, other principles of operation are also utilized [7]. The principle of the operation of these devices is based on multicomponent measurements of the sensor arrays coupled with various pattern recognition methods. The procedures of analyzing responses of the sensor systems rely on the application of statistical and mathematical methods and they demand sophisticated methods originating from chemometrics, a subdiscipline of chemistry [8].

PCA is the most common and versatile method to display electronic tongue and electronic nose measurements. It decomposes the data matrix into a new set of uncorrelated variables (Principal Components), which means that it finds new directions in the pattern space, so that they explain the maximum amount of variance in the data set. These new variables may be used as inputs for more complex classifiers, e.g. ANN. In contrast to PCA, PLS-DA is a supervised method, which models the relationship between two matrices, i.e., the data set obtained from sensor array measurements and class affiliation matrix (target matrix composed of vector with true class affiliations). PLS-DA determines a set of latent variables corresponding to principal components in PCA, but explaining as much as possible of the covariance between the two matrices (PLS-DA scores). This is a generalization of multiple linear regression; it can analyze more noisy and uncompleted data and it is able to manage with multicolinearity problem, which often occurs in sensor array measurements [9]. The output of PLS-DA is the score matrix (PLS-DA scores) that can be plotted similarly as in

* Corresponding author. Tel.: +48 22 6607873; fax: +48 22 6605631.
E-mail address: pciosek@ch.pw.edu.pl (P. Ciosek).

0039-9140/$ – see front matter © 2005 Elsevier B.V. All rights reserved.
doi:10.1016/j.talanta.2005.03.006

PCA, and predictor matrix (Ypred), which estimates class affiliation. The comparison of particular vectors of predictor matrix with respondent vectors of target matrix shows correctness or incorrectness of particular sample classification. When comparisons of all vectors are performed, percent of correct classifications is obtained:

$$\%CC = \frac{N_c}{N_c + N_{nc}} \times 100\%$$

where N_c is the number of correct classifications and N_{nc} is the number of incorrect classifications.

ANNs, a powerful tool for non-linear approximations, are widely used in artificial senses data analysis because of their ability to imitate human brain behavior learning the solution of problems from the data avoiding the necessity of any modeling. Among numerous network architectures, back-propagation neural networks have been frequently used [10]. The output of ANN is predictor matrix (Ypred), which after comparison with target matrix gives percent of correct classifications (the same procedure as for PLS-DA).

PCA, PLS-DA and ANN may be combined in order to enhance classification capability. In this paper, the comparison of this method for the classification of commercial brands of orange juices has been presented. Measurements were performed by an electronic tongue developed at Warsaw University of Technology [11]. This work presents also, a new strategy for classification based on the use of a supervised feature extraction in order to enhance the classification capability of the system.

2. Sensor array

The sensor array used in the experiment was formed by 16 ion-selective electrodes (IS 561, Philips) and one standard pH electrode (Mettler Toledo InLab 407 connected to pH-meter Mettler Delta 350), which creates an array of 17 sensors [11]. The system was composed of eight types of electrodes: four types of classical ion-selective electrodes with enhanced selectivity towards particular ionic species present in the sample, and four electrodes with enhanced cross-sensitivity. Two electrodes of the same type for each membrane composition were prepared. The membranes for ISEs preparation contained appropriate ionophore, lipophilic salt, 61 wt.% plasticizer, and 31–33 wt.% high-molecular-weight PVC (Table 1). Membrane components were supplied by Fluka Chemie AG, i.e., TPPClMn (chloride ionophore I), ETH 6010 (carbonate ionophore I), valinomycin (potassium ionophore I), ionophore X (sodium ionophore X), nonactin (ammonium ionophore I), ETH 129 (calcium ionophore II), TDMAC (tridodecylmethylammonium chloride), TDAB (tetrakis(decyl)ammonium bromide), KTFPB (potassium tetrakis [3,5-bis(trifluoromethyl)phenyl]borate), KTPClPB (potassium tetrakis (4-chlorophenyl)borate), *o*-NPOE (2-nitrophenyl octyl ether), DOS (bis(2-ethylhexyl)sebacate) and BBPA (bis(1-butylpentyl)adipate). Fluoride ionophore (uranyl salophene derivative) was synthesized in Laboratory of SMCT, MESA+ Research Institute, University of Twente [12].

3. Measurements

All measurements were carried out at room temperature (20 °C) using a multiplexer (EMF 16 Interface, Lawson Labs Inc., accuracy of measurement −0.1 mV) with cells of the following type: Ag, AgCl; KCl 1 M/CH_3COOLi 1 M/sample solution//membrane//internal filling solution; AgCl, Ag. This system enables to perform measurements of liquids as conventional direct potentiometry without any sample pretreatment.

Five brands of orange juice commercialized in Poland: Cappy, Fortuna, Tarczyn, Clippo, Hortex, were measured. For each brand of juice, samples with different manufacture date (from various manufacture lots) were used. The learning set was constructed from the data obtained by measuring samples from two different manufacture lots and the test set comprised samples originated from a third different manufacture lot. The independence of learning and testing set was in this way provided. For each brand and each lot, six samples were measured.

The measurement of the electrodes signals in each sample lasted 15 min in 5 s intervals. After reaching steady-state

Table 1
Components used for sensor membranes preparation

Electrode number	Electrode type	Plasticizer	Lipophilic salt	Ionophore
1, 2	Ca^{2+}	*o*-NPOE	KTFPB	2 wt.% ETH 1001
3, 4	NH_4^+	BPPA	KTPClPB	2 wt.% nonactine
5, 6	Na^+/K^+	*o*-NPOE		5.15 wt.% ionophore X 0.2 wt.% valinomycin
7, 8	Cl^-	*o*-NPOE		1 wt.% TPPClMn
9, 10	HCO_3^-	*o*-NPOE	TDMAC	1 wt.% ETH 6010
11, 12	"Cation-selective"	DOS	KTPFPB	–
13, 14	$F^-/H_2PO_4^-$	*o*-NPOE	TDAB	1.5 wt.% ionophore $H_2PO_4^-$; 0.05 wt.% ionophore F^-
15, 16	"Anion-selective"	*o*-NPOE	TDMAC	–

responses, the last 10 values of the electrodes potentials were averaged in pairs, i.e., from every type of sensors 10 outputs formed inputs for further data analysis. The learning set comprised of: 2 (lots) × 5 (brands) × 6 (samples) × 10 (measurements) = 600 cases belonging to five classes (brands), and the testing set 1 (lot) × 5 (brands) × 6 (samples) × 10 (measurements) = 300 cases.

4. Data analysis

Mean values of the responses of the electrodes of the same type were calculated for each sample to form inputs for further data analysis process. To remove scale effects, sensor outputs were autoscaled (*z* transformation) [13].

All pattern-recognition techniques are based on learning by example, i.e., having a set of feature patterns of known class (learning set); the classifier system is learned to give corresponding class membership responses. Each sample was characterized by input vector (features obtained by measuring the sample) and five-dimension vector (target vector), which should appear at output of the classifier to properly classify the sample. Five brands of juices were used, so the simplest way of creating target vector was to put in order successive samples: (1) Cappy, (2) Fortuna, (3) Tarczyn, (4) Clippo, (5) Hortex). Each of five outputs marked successive number of

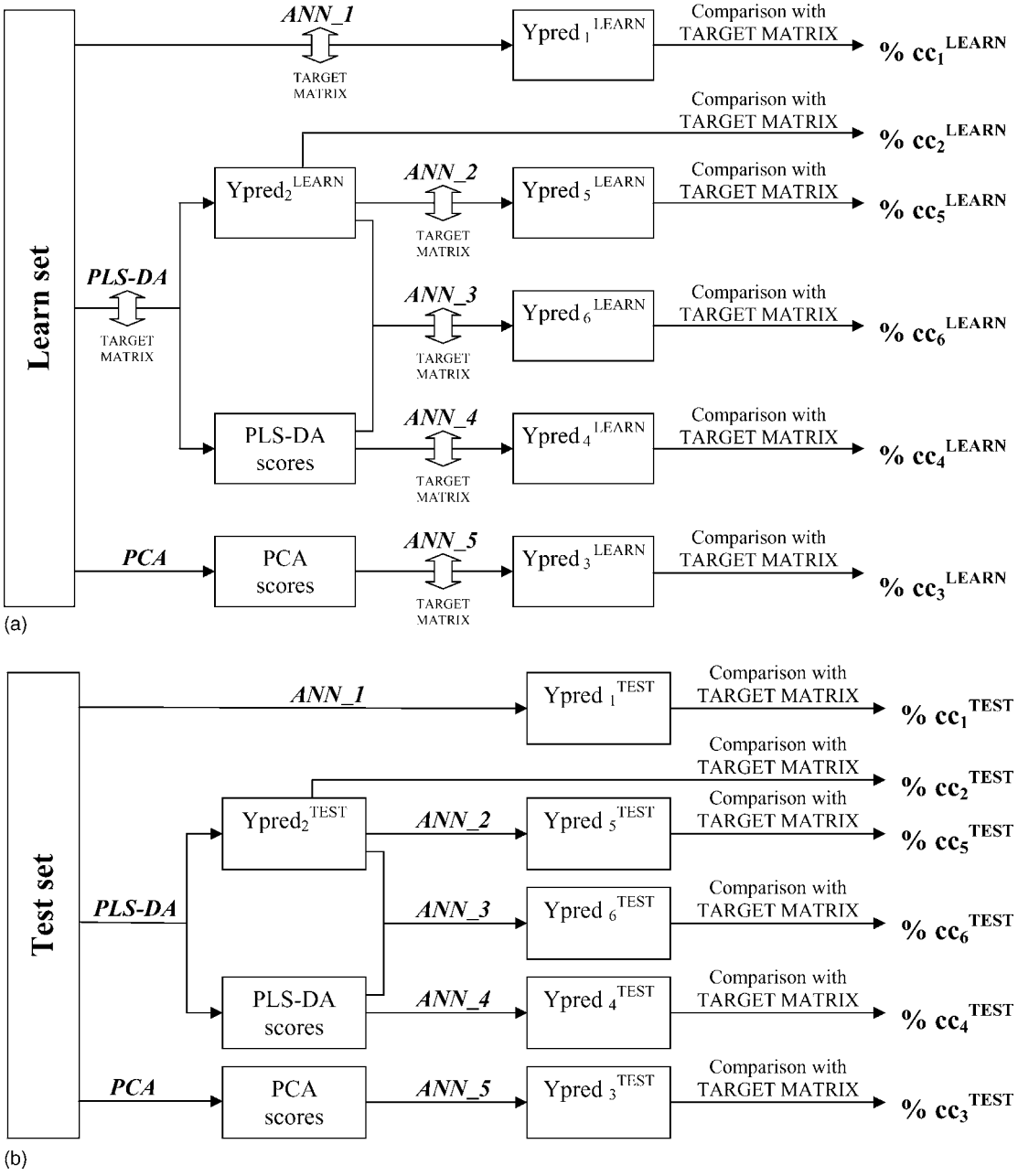

Fig. 1. Processing methods presented in the article: (a) establishing of the models (fitting), (b) validation of the models with independent test samples.

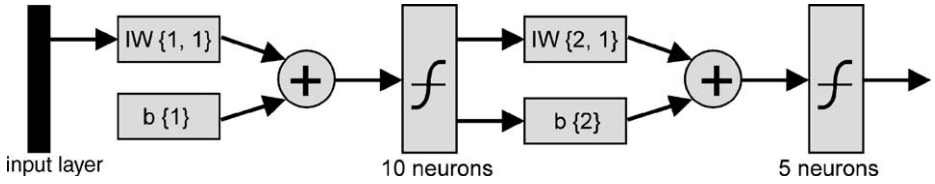

Fig. 2. Type of architecture of the neural network.

the sample. If, for example at fourth output, +1 appeared and 0 at the rest of outputs ("one-of-many" code [11,14]), then it meant that the sample's number was 4 (Clippo). Process of learning involved adjusting value of weights and biases of each neuron (IW {1, 1}, b {1}, LW {2, 1}, b {2}, Fig. 2) or finding LVs (PLS) in order to provide desired outputs corresponding to a determined input (see Fig. 1a). When satisfactory level of error for the object from learning set is obtained, class membership of unknown sample can be estimated. When obtained classifications are compared with true class membership (Comparison with TARGET MATRIX, see Fig. 1b), percent of correct classifications is calculated.

Two kinds of data treatment were carried out: direct processing and two-stage processing (Fig. 1). In direct processing, two methods were used: ANN and PLS-DA (resulted in $\%cc_1$ and $\%cc_2$, respectively). In the two-stage processing, calculations based on four methods were performed. One of them involved PCA as a feature extraction phase for ANN processing (resulted in $\%cc_3$). This combination has been utilized by several authors [2,6,11,15]. Although this approach has shown important advantages, it is not the optimal choice. PCA, used as pre-processor, puts in evidence the most correlated part of a data set and sometimes this point of view is not optimal for classification purposes. It is possible that the portion of data interesting for classification does not coincide with that of maximum correlation, or in PCA language that carrying the maximum portion of variance. Nonetheless, the advantage of PCA is in the fact that it removes correlation between variables making the application of complex classifiers easier like neural networks.

On the contrary, it is possible to refine the pre-processing extracting features that are actually correlated with the solution of the problem. This opportunity is offered by PLS-DA. Indeed, this method can be illustrated as a PCA where the scores are further rotated in order to maximize the correlation with scores of the target matrix.

Although it has been developed as a regression method, PLS can be utilized to solve classification problems encoding in the target matrix the class membership of the measured samples (PLS-DA). The usual "one-of-many" code has been here utilized. The same coding has also been used to train ANN.

Methods based on the assumption that the use of a supervised method also in the feature extraction phase could enhance classification capability of the system, has not been presented in the literature so far. As a result of PLS-DA two matrices are obtained: "scores" and "ypred" (matrix of predicted affiliation to the particular class predictor matrix). Both of them can be used separately (resulted in $\%cc_4$ and $\%cc_5$) or together (resulted in $\%cc_6$) for feature extraction phase for further ANN processing. In all experiments, the same architecture of back-propagation neural network (Fig. 2), i.e., the sigmoid transfer function and gradient descent algorithm (learning rate = 0.5, momentum coefficient = 0.8) to adjust weights and biases in the network was used. The networks contained 10 neurons in hidden layer and five in output layer

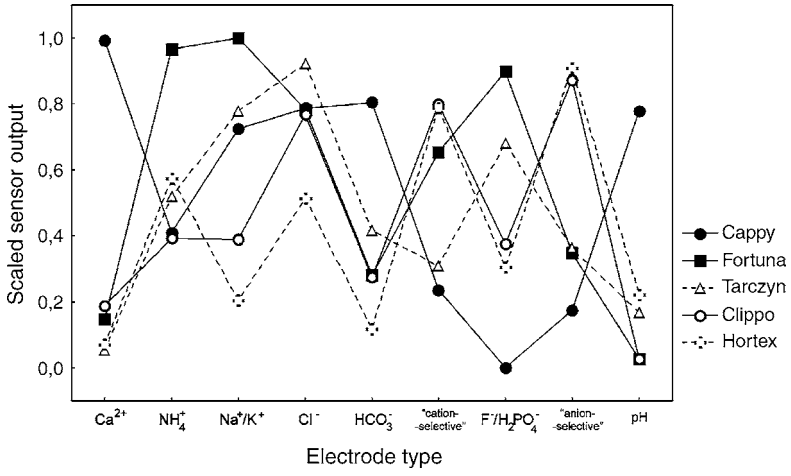

Fig. 3. Scaled responses for orange juices (all brands).

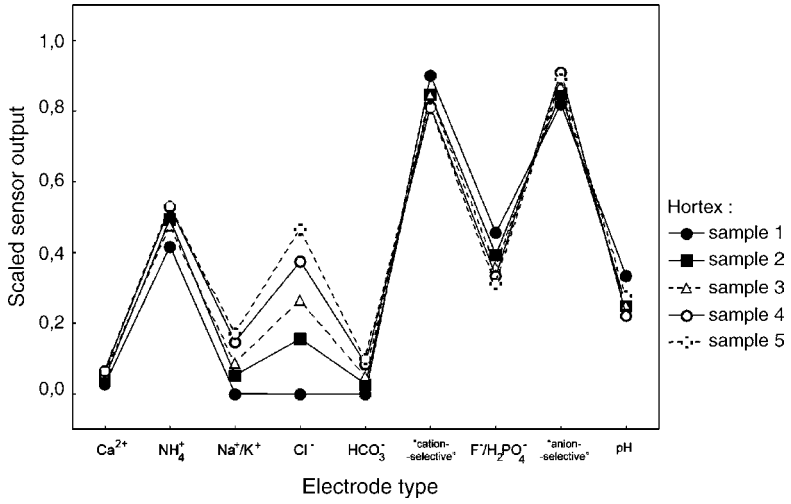

Fig. 4. Scaled responses for orange juices (one brand).

(corresponding to five brands of juice). All PCs and LVs were considered, i.e., nine. The number of neurons in input layer was dependent from the number of columns in the data matrix (nine columns in all cases besides PLS-DA (ypred) + ANN model—five neurons, and PLS-DA (scores + ypred) + ANN model—14 neurons). The data processing was realized in MatLab (The MathWorks Inc., Natick, USA).

5. Results and discussion

Scaling of electrode responses was performed in order to visualize sensor array outputs (Figs. 3 and 4). All the brands were easily distinguished from each other—their pattern of responses were evidently different (Fig. 3). Similarity of pattern responses of the same brand samples is presented in Fig. 4. Responses of the majority of the electrodes were almost the same; however, two electrodes (Cl^-, Na^+/K^+-selective) exhibited different signals for the same brand samples. This effect is observed for the electrodes possessing the worst discrimination capability, i.e., the sensors for which the range of the potentials measured in all brands tested is the narrowest. Therefore, the scaled responses of these sensors measured in the same brand sample are characterized by the greatest signal variability in the [0,1] interval.

Building up learning and testing set from the measurements of samples originating from various manufacture lots was undertaken in order to avoid overfitting of the classification model. The authors observed that the data set from the samples from one manufacture lot usually produce overfitted models, which means that samples with the same composition (the same manufacture lot) are recognized with 100% accu-

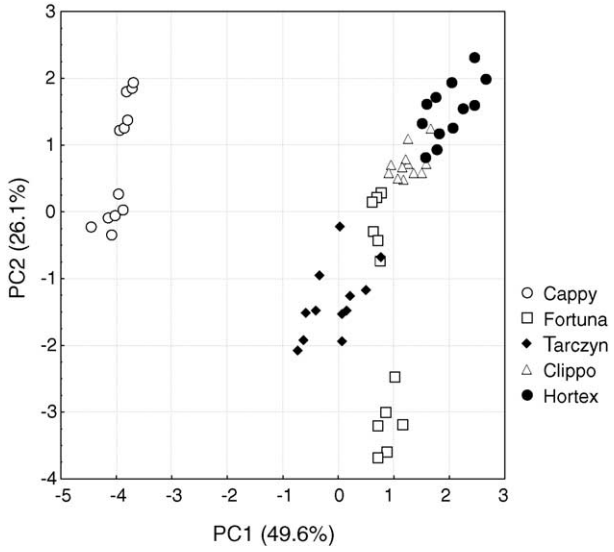

Fig. 5. PCA plot of juice measurements (for each brand samples from two manufacture lots were presented).

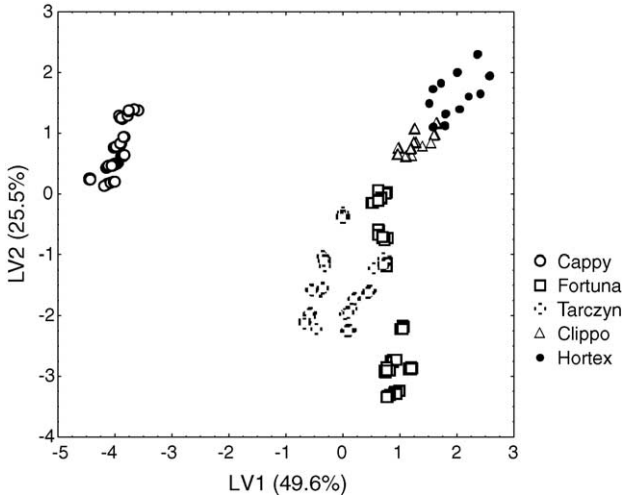

Fig. 6. PLS plot of juice measurements (for each brand samples from two manufacture lots were presented).

Table 2
Results of classification

Model	Model performance for the learning set (fitting)	Model performance for the testing set (validation)
Direct processing		
ANN	$\%cCC_1^{LEARN} = 20.0$[a]	$\%cCC_1^{TEST} = 20.0$
PLS-DA	$\%cCC_2^{LEARN} = 100.0$	$\%cCC_2^{TEST} = 93.0$
Two-stage processing		
PCA + ANN	$\%cCC_3^{LEARN} = 100.0$	$\%cCC_3^{TEST} = 92.0$
PLS-DA (scores) + ANN	$\%cCC_4^{LEARN} = 100.0$	$\%cCC_4^{TEST} = 70.0$
PLS-DA (ypred) + ANN	$\%cCC_5^{LEARN} = 100.0$	$\%cCC_5^{TEST} = 90.0$
PLS-DA (scores + ypred) + ANN	$\%cCC_6^{LEARN} = 100.0$	$\%cCC_6^{TEST} = 100.0$

[a] Establishing of the model failed.

racy, when those with different manufacture date are hardly or even completely incorrectly recognized by the system. It is probably due to the fact that in this case, the classifier is too complex; it models noise in the data set and it fails to capture true data structure. In order to improve the generalization performance and thus working out the true working condition, the data obtained by measuring samples originating from different manufacture lots was used to construct the classification model. The testing set was independent of the learning one (different manufacture lot), and thus, capturing of true data structure of the model was possible to be checked.

PCA and PLS-DA were used to extract information from multicomponent measurements and to remove redundant data. They also allow the visualization of the majority of significant data in two or three-dimensional spaces. Respective clustering of measurement data coming from various manufacturers of orange juice was visible both on PCA and PLS-DA plots (Figs. 5 and 6, respectively).

In the case of direct processing, satisfactory results were achieved by PLS-DA (Table 2, $\%cc_2 = 93$). This was due to the fact that this classification method besides removing redundant data and giving uncorrelated features, provides features that are at their maximum extent correlated with the classification objective. ANN without any feature extraction procedure was not able to give reasonable results—its classification accuracy was equal to casual classification (probability of assigning the sample to one of five classes was equal 20%, the size of each class was the same). ANN needed to perform feature extraction phase, at other times it could not find true data structure.

In the case of two-stage data processing, noteworthy results were achieved by the sequence PCA and ANN ($\%cc_3 = 92$). PLS-DA as a tool for extraction of significant data was able to provide satisfactory results only when two matrices, “scores” and ”ypred” (prediction matrix), were used. It has to be noted that these matrices represents the described sensors data (scores) and the predicted class membership (ypred). By using both of that, all the potentialities of PLS-DA are taken into account as variables to be treated by the ANN. PLS output matrix (ypred) is built up of vectors, in which the highest value determines class membership. For example vector [0.9, 0.1, 0.9001, 0, 0.2] indicates that sample belongs to third brand. However, this assumption is problematic, since the sample is also very similar to the first brand. In such cases, additional information processed directly by ANN can be considered and appropriate choice between these two brands can be done—that is why the best performance was obtained when two matrices, ypred and scores, were processed by ANN. The role of ANN is then to combine nonlinearly all the information that is necessary for classification and that it is provided by PLS-DA. This method resulted in 100% classification capability (%cc6, Table 2).

6. Summary

The comparison of two methods, direct and two-stage processing, for data obtained by measuring samples of orange juices was presented in this paper. An array of ion-selective electrodes containing two kinds of sensors, one with enhanced selectivity towards ionic species present in the sample, and the other, cross-sensitive, was able to discriminate between various brands of orange juices originating from various manufacture lots. Classification ability of various procedures based on PCA, PLS-DA and ANN was presented.

A new method of sensor array data analysis, based on the fact that supervised feature extraction could enhance classification ability of multi-sensor systems, was introduced. Verification of the presented method resulted in a complete identification of unknown samples.

Acknowledgment

This paper is financed by The Foundation for Polish Science within frame of professor fellowship.

References

[1] K. Persaud, G.H. Dodd, Nature 299 (1982) 352–355.
[2] H.T. Nagle, S.S. Schiffman, R. Gutierrez-Osuna, IEEE Spectrum 7 (1998) 22–34.
[3] M. Otto, J.D.R. Thomas, Anal. Chem. 57 (1985) 2647–2651.
[4] K. Toko, Mater. Sci. Eng., C 4 (1996) 69–82.
[5] F. Winquist, P. Wide, I. Lundstrom, Anal. Chim. Acta 357 (1997) 21–31.

[6] A. Legin, A.M. Rudnitskaya, Y. Vlasov, C. Di Natale, F.A.M. Davide, A. D'Amico, Sens. Actuators B 44 (1997) 291–296.
[7] G. Sehra, M. Cole, J.W. Gardner, Sens. Actuators B 103 (2004) 233–239.
[8] P.K. Hopke, Anal. Chim. Acta 500 (2003) 365–377.
[9] S. Wold, M. Sjostrom, L. Eriksson, Chemom. Intell. Lab. Syst. 58 (2001) 109–130.
[10] P. Daponte, D. Grimaldi, Measurement 23 (1998) 93–115.
[11] P. Ciosek, Z. Brzózka, W. Wróblewski, Sens. Actuators B 103 (2004) 76–83.
[12] K. Wojciechowski, W. Wróblewski, J. Przygórzewska, G. Rokicki, Z. Brzózka, Chem. Anal. (Warsaw Pol.) 47 (2002) 335–346.
[13] D.L. Massart, B.G.M. Vandeginste, S.N. Deming, Y. Michotte, L. Kaufman, Chemometrics: a Textbook, Elsevier, Amsterdam, 1998.
[14] P.J. O'Riordan, C.M. Delahunty, Int. Dairy J. 13 (2003) 355–370.
[15] P. Ciosek, E. Augustyniak, W. Wróblewski, Analyst 129 (2004) 639–644.