Design and Implementation of an Electronic Nose Based on Power-Law Model

Walaa Mohammed H. Khalaf

Lecturer
Computer and Software Engineering Dep
Almustansirya University
walaakhalaf@yahoo.com

Giancarlo Raiconi

Associate Professor
Department of Mathematics and Informatics Universityof Salerno, 84081 Baronissi Salerno, Italy.

Manlio Gaudioso

Professor
Department of Electronics, Informaticsand Sistems University of Calabria, 87036 Rende (CS), Italy

Abstract

Electronic noses are intelligent systems having the ability to identify and quantify different gases. They are capable to handle large data sets and to perform quite complex signal processing by employing commercial sensors. In this paper we examine four different gases (ethanol, methanol, acetone and benzene). The data are acquired and processed by a prototype multi-sensor monitoring system (electronic nose) which has been designed at University of Calabria. It includes five gas sensors plus two auxiliary sensors, for temperature and humidity, respectively. The system is equipped by an appropriate interface to a PC for visualizing and analyzing the data.

We introduce an analytical power law model of sensor response. Least square regression is performed to estimate model parameters. The model is then used for both gas classification and concentration estimate. The results of the numerical experiments on 172 gas samples are reported.

1. Introduction

In recent years, there has been a lot of interest in investigating the possibility of using machine learning techniques in the framework of volatile organic compounds (VOCs) [1]. The aim is to classify the different types of gases and also, to estimate their concentrations. The subject is relevant from practical point of view as the possible toxic nature of VOCs at high

concentration levels may provoke serious dangers and it is at the basis of several allergic pathologies, skin and lung diseases [2].

An Electronic Nose (E-nose) is a system consisting of an array of sensors which are specifically designed to respond to the surrounding odorants; they use chemical polymer or metal oxide semiconductor (see e.g. [3, 4, 5]). The output of an E-nose is converted to an electronic signal which in turn is conveyed to a computer for applying various classification approaches.

Chemoresistive SnO_2 gas sensors are very sensitive but not selective [6]; their cost is low and they operate on the principle that their resistance changes in the presence of oxidizing or gas reducing processes.

In many practical situations is not known in advance the type of gas the sensor may be exposed to. Consequently a relevant problem is not only to estimate gas concentration, but also to identify the gas type from a possible set. Such task cannot be performed by just a single sensor [13]. In fact cannot be associated the sensor response, in terms of resistance variation, to a specific gas type.

To detect both gas type and concentration we use more than one gas sensor, because a single gas sensor would give the same response (resistance value or, equivalently, voltage) to different concentrations of different gases. On the contrary, appropriate combination of the response curves of several sensors can provide the requested answers.

In fact, E-nose techniques are developed in order to classify the target gases. This can be considered as a pattern classification problem [5][7]. The classification is usually performed in two stages: training and validation. Another specific requirement for E-nose is the gas concentration. For example, in the international shuttle space (ISS), the spacecraft maximum allowable concentration (SMAC) is defined as a guideline; the air quality is continuously monitored to comply with the SMAC, so that the astronauts can live in a healthy and safe environment [8].

In this work we consider a supervised learning technique to solve the classification problem which benefits from the introduction of an explicit model of the distribution of the data.

This paper is organized as follows. In Section 2 we present an overview of our electronic nose system, while in section 3 we introduce the mathematical model. In section 4 description of the experimental data is provided, while the experimental results of our implementation are shown in section 5. Some concluding remarks are in section 6.

2. Electronic Nose

An electronic nose combines an array of gas sensors, whose response constitutes an odor pattern, and a machine learning system having the ability to discriminate among different odorant types as shown in Fig. 1. Our electronic nose system contains an array of five tin oxide based

Taguchi gas sensors, from Figaro Engineering, plus a temperature sensor and a humidity sensor [10].

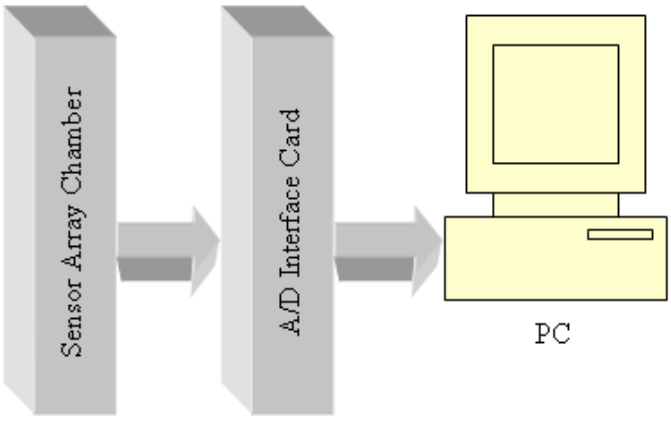

Fig. 1. General Scheme of Electronic Nose System.

The sensing element for each gas sensor is a tin dioxide (SnO_2) semiconductor. In particular three of them are of the TGS–822 type, each one being supplied with a specific heater voltage (5.0 V, 4.8 V, and 4.6 V, respectively), one of the TGS–813 type, and the last one is of the TGS–2600 type. The two auxiliary sensors are the temperature sensor (LM35 from National Semiconductor Corporation), and the humidity sensor (HIH-3610 from Honeywell) [10]. Their presence is motivated by the strong effect that both humidity and temperature changes have on the behavior of most gas sensors. The gas sensors and the auxiliary sensors are put inside a box with effective volume of 3000cc. Inside the box we put also a fan to let the solvent drops evaporate easily.

All sensors are connected with a multifunction board (NI DAQPad-6015), which is used in our system as an interfacing between the BOX and the PC. NI-DAQmx provides an interface to our LabWindows/CVI that we used inside the PC.

LabWindows package is a programming tool running on a PC Pentium 4 type. The integrated LabWindows/CVI environment features code generation tools and prototyping utilities for fast and easy C code development. It offers a unique, interactive ANSI C approach that delivers access to the full power of C Language. Because LabWindows/CVI is a programming environment for developing measurement applications, it includes a large set of run-time libraries for instrument control, data acquisition, and analysis.

A machine learning system is divided into two parts (modes), training mode and testing mode; in training mode the system finds the optimal parameters for both classification and

concentration estimate. In testing mode the system uses parameters calculated during the training mode to analyze a gas sample of unknown type and concentration. The system scheme is represented in Fig. 2.

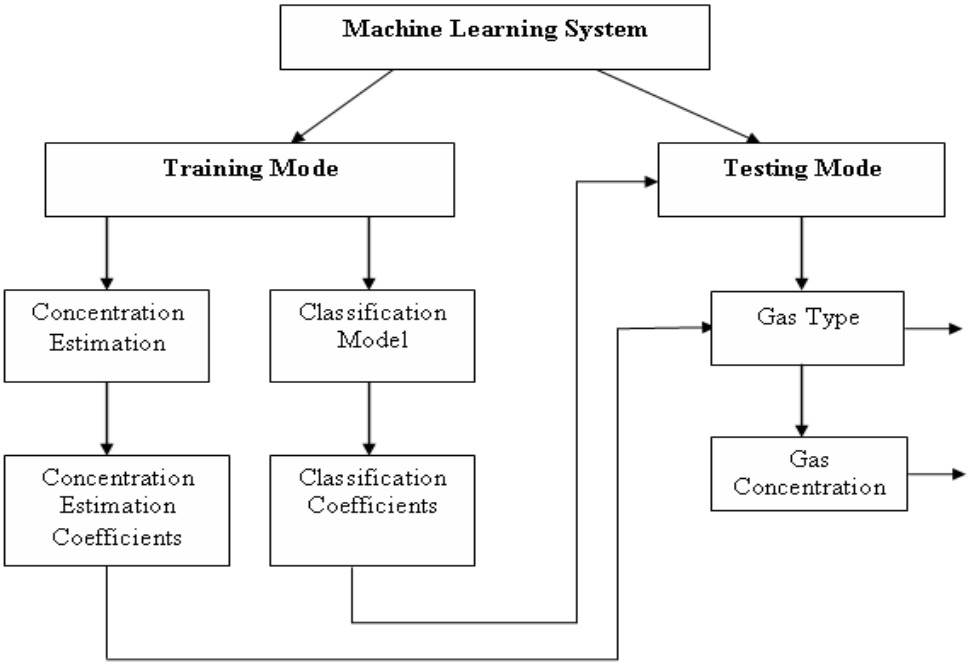

Fig. 2. Scheme of the machine learning system.

3. The classification model

Now we formalize our approach. For a certain gas index i (i = 1, 2, …, n_G), where n_G represents the number of gases under investigation, the concentration (c_i), as a function of the measures obtained by several sensors, is described as a curve in a n_S dimension space (n_S represents the number of sensors). The parametric equations are:

$$v_j = X^i_j(c_i), \qquad j = 1, 2, \ldots, n_S \qquad \ldots\ldots\ldots(1)$$

where v_j is the response of sensor j, X^i_j is the response function to the gas i of sensor j. If we assume that a given number of samples N_i , i=1,…, n_G, of gas i. is available , we write:

$$z_{jk} = X^i_j(c_{ik}) + w_{ik}\,, \qquad k = 1, 2, \ldots, N_i, \quad j = 1, 2, \ldots, n_S \qquad \ldots\ldots\ldots(2)$$

where z_{jk} represents the response of sensor j for the k *th* sample of gas i, w_{ik} represents the measurement error, It is then possible to build a data fitting strategy using the least squares approach and a parametric model of the functions:

$$X_j^i(c_i) = X(c_i, \theta_j^i), \qquad \theta_j^i \in \Re^p \qquad \text{.....(3)}$$

where θ represents the equations parameters. From qualitative evaluation of the response curves, it appears reasonable to adopt the same model for all gases and sensors. Once the parameters have been estimated on the basis of sampling data, the parametric curve equations, for each gas i, are:

$$x^i = \begin{bmatrix} X(c, \hat{\theta}_1^i) \\ X(c, \hat{\theta}_2^i) \\ \cdots\cdots\cdots\cdots \\ X(c, \hat{\theta}_{n_s}^i) \end{bmatrix}, \qquad c_{i,\min} \le c \le c_{i,\max} \qquad \text{.....(4)}$$

Observe that two of such curves in the space R^n_s may intersect, only if there exists a pair of gases i and l, and two concentration values c_i and c_l such that:

$$X(c_i, \hat{\theta}_j^i) = X(c_l, \hat{\theta}_j^l), \qquad j = 1,2,\ldots,n_S \qquad \text{.....(5)}$$

The construction of the curves represents the phase where our classifier is being built. It can be used, as a classification tool, as follows.

Suppose that a newly generated set of measures $z = [z_1, z_2, \ldots, z_{ns}]^T$ carried out on an unknown gas, at unknown concentration is available. The most natural way to obtain a classification of the new sample is to identify the gas type with associated the minimum distance between the sample and the curve, that is by detecting the one which provides the minimum distance:

$$dist(z, x_i) = \min_{x \in x_i} \|z - x\|^2 \qquad \text{..... (6)}$$

Summing up, the classification-evaluation problem is divided into two steps:

1. Construction of the model (for different gases).
2. Solution of minimum distance optimization problem.

4. Implementation

By considering the data, we can find typical curves representing the response of one gas sensor *j* with respect to one gas type *i* at different concentrations). A typical behavior is given in Fig.3. It represents the response, in terms of output voltage (expressed in Volts) of sensor TGS-822 for Acetone gas type at the concentrations given in abscissa, which are measured in parts-per-million (ppm).

Instead of using the voltage as the output variable of the gas sensors, we suggest to use directly the resistance response. Such value can be obtained through the transformation:

$$R_S = ((V_c / V_o) - 1) * R_L \qquad \text{.....(7)}$$

where R_S is the sensor resistance response, V_c is the source voltage (V_c=5.0V, in our experiments), V_o is the output voltage (sensor voltage response), and R_L is the load resistance. If we plot the experimental curve between gas concentration in parts-per-million (ppm) and the sensor resistance response (in Ohm) we will obtain the following curve (as shown in Fig. 4).

The qualitative behavior exhibited by the empirical results, suggests to adopt a curve fitting model of the power-law type for each sensor j and for each gas *i*, that is:

$$R_j(c_i) = \zeta_j^i(c_i, \beta_j^i, A_j^i) = A_j^i c_i^{\beta_j^i} \qquad \text{..... (8)}$$

where we have made explicit the meaning of the parameters θ.

Calculation of the model parameters has been performed by using the MATLAB toolbox function **ezyfit**.

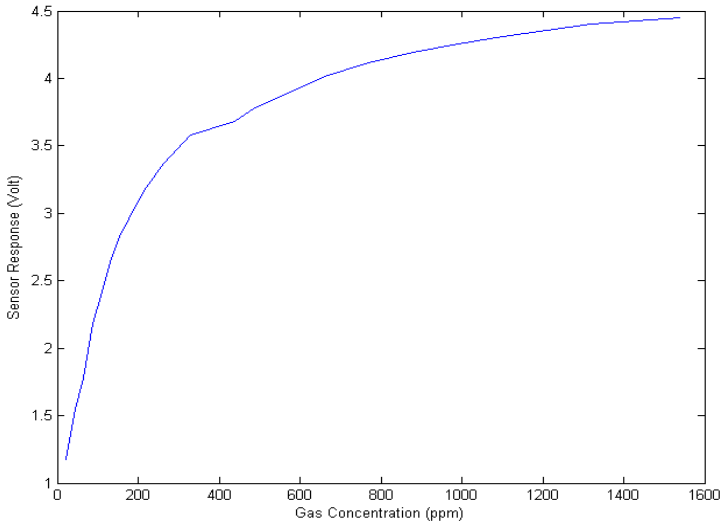

Fig. 3 . Sensor Response of Acetone Gas.

Once the (2 * n_G * n_S) parameters for A and β have been obtained, we have implemented the detection phase. For any gas sample $z \in R^{n_S}$, where z is the collection of the measures obtained from the different sensors exposed at any gas at any concentration, we have classified the gas (that is guessed the gas index *i*) and estimated the concentration c_{GAS} by the following formulae:

$$i_{GAS} = \arg\min_{i} \left\{ \min_{c} \sum_{j=1}^{n_S} (z_j - x(c, \theta_j^i))^2 \right\} \qquad \ldots\ldots\ldots (9)$$

and the concentration as

$$c_{GAS} = \arg\min_{c} \sum_{j=1}^{n_S} (z_j - x(c, \theta_j^{i_{GAS}}))^2 \qquad \ldots\ldots\ldots (10)$$

5. Results

We applied the method of leave–one-out cross-validation on our data set which contains 4 different types of gases which are Acetone, Ethanol, Methanol, and Benzene for different range of concentrations (172 gas samples), each time we considered one sample as test sample and the

remaining samples as training set. For median-sized problems, cross validation might be the most reliable way for parameter selection. First, the training data is partitioned into several folds. Sequentially a fold is considered as the validation set and the rest are for training. The average of accuracy on predicting the validation sets is the cross validation accuracy [9]. In particular the leave-one-out cross validation scheme consists of defining folds which are singletons, i.e. each of them is constituted by just one sample. By applying **172** cross validations on our data set we got **98.8372%** as correct gas classification rate.

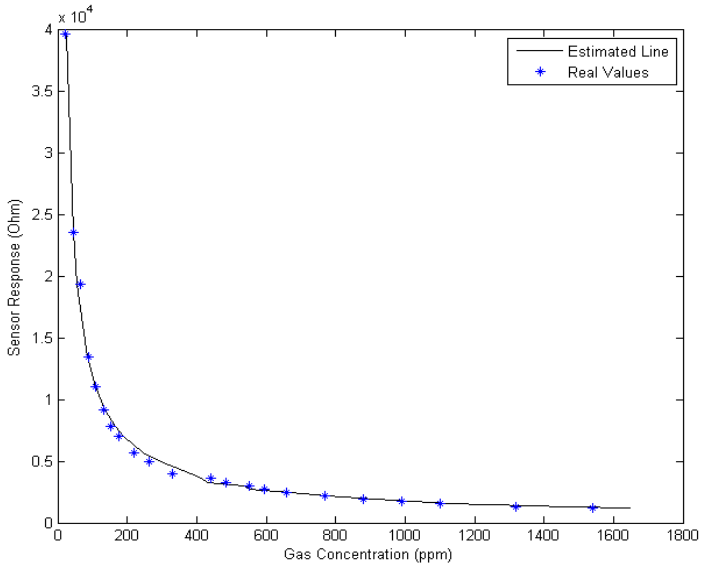

Fig. 4. Sensor Response (in Ohm) the real values and estimated curve.

As for concentration estimate, we used the correlation coefficient as a measure of the concentration estimation quality [12].

$$C.C = \frac{\sum_{i=1}^{n} X_i \hat{X}_i - \frac{\sum_{i=1}^{n} X_i \sum_{i=1}^{n} \hat{X}_i}{n}}{\sqrt{\left(\sum_{i=1}^{n} X_i^2 - \frac{(\sum_{i=1}^{n} X_i)^2}{n}\right)\left(\sum_{i=1}^{n} \hat{X}_i^2 - \frac{(\sum_{i=1}^{n} \hat{X}_i)^2}{n}\right)}} \quad \text{.....(11)}$$

where *C.C* is the correlation coefficient, *X* are the actual values, ^*X* are the predicted values, and *n* is the number of data points.

Note that such coefficient assumes a value close to zero when the estimation rate is poor, whereas it is equal to one if we have a perfect fit. The values we got are: for Acetone we got the correlation coefficient equal to **0.9969**, for Ethanol **0.9749**, for Methanol **0.9982**, and for Benzene **0.9867**.

Our results demonstrate that the system has a good ability to identify the gas type as well as to quantify its concentration.

References

1. Robi Polikar , Ruth Shinar, Vasant Honavar, Lalita Udpa, Marc D. Porter Detection and Identification of Odorants using an Electronic Nose, IEEE, 2001, Page 3137-3140
2. Casalinuovo, I.; Pierro, D. **Application of Electronic Noses for Disease Diagnosis and Food Spoilage Detection**. Sensors, 2006, 6, 1428-1439.
3. J. W. Gardner and P. Barlett, "**A brief history of electronic noses**," Sensors and Actuators B, vol. 18, pp. 211–220, 1994.
4. T. C. Pearce and M. A. Sanchez-Montanes, "**Chemical sensor array optimization: Geometric and information theoretic approaches**," in Handbook of Machine Olfaction, T. C. Pearce, S. S. Schiffman, H. T. Nagle, and J. W. Gardner, Eds. Weinheim, Germany:Wiley-VCH, 2002, pp. 347–376.
5. Dongliang Huang, and Henry Leung,, **Simultaneous Classification and Concentration Estimation for Electronic Nose**, IEEE SENSORS JOURNAL, VOL. 7, NO. 5, MAY 2007, pp. 825-834.
6. R. Ionescu. M. Pardo and G. Sberveglieri. **Classification of electronic nose data with support vector machines**. Sensors and Actuators B, 48:392–394, 1998.
7. M. Pardo and G. Sberveglieri, "**Learning from data: A tutorial with emphasis on modern pattern recognition methods**," IEEE Sensors J., vol. 2, no. 3, pp. 203–217, Jun. 2002.
8. M. A. Ryan et al., "**Monitoring space shuttle air quality using the jet propulsion laboratory electronic nose**," IEEE Sensors J., vol. 4, no. 3, pp. 337–347, Jun. 2004.

9. Gallant, S.I. Neural Network Learning and Expert Systems. MIT press: Cambridge,
Massachusetts, 1993.
10. Khalaf Walaa, Pace Calogero, Gaudioso Manlio, "**Least Square Regression Method for Estimating Gas Concentration in an Electronic Nose System**". Sensors, 2009, Volume 9, no. 3, pp. 1678-1691.
11. Yang Jianhua, Hou Hong, Wang Lei. "**Electronic nose system based on integrated gas sensor aray,**"Journal of Transducer Technology, Vol. 33(8), 21-23, 2003.
12. Penza, M.; Cassano, G.; and Tortorella, F. Identification and Quantification of Individual Volatile Organic Compounds in a Binary Mixture by Saw Multisensor Array and Pattern Recognition Analysis. Measurment Science and Technology, 2002, 13, 846-858.
13. Bartlett, P.; and Gardner J. Odour Sensors for an Electronic Nose, in Sensors and Sensory Systems for an Electronic Nose. NATO ASI Series E: Applied Science, 1992, 212, 197-216.