Independent Component Analysis for Pattern Recognition of Gas Sensor Array Measurement Data

Kai Song[1], Qi Wang[1], Taehwan Kim[2], and Keunsung Bae[2]
[1] School of Electrical Engineering and Automation, Harbin Institute of Technology
Harbin 150001, China
[2] School of Electrical Engineering and Computer Science, Kyungpook National University
Daegu 702-701, Korea
E-mail: [1]kaisong@hit.edu.cn, [2]ksbae@knu.ac.kr

Abstract: Commercial metal oxide semiconductor gas sensor array system is generally subject to significant cross-sensitivity due to poor selectivity for different gases. This crosstalk greatly affects the accuracy of gas recognition and concentration measurement. In this paper, we explore the independent component analysis as a novel modeling technique to process the gas sensor array response data for quantitative and qualitative analysis. The process consists of two stages: first the type of gas is classified from the independent components analysis, then gas concentration is computed using the scatter plot of adjacent independent components. Simulation results demonstrate that the ICA-based method show good performance on the qualitative and quantitative gas component analysis.

1. Introduction

Each sensor in commercial gas sensor array does not behave as an independent sensor, rather it responds to a variety of gases with varying sensitivity. Thus a sensor output may contain the information of gases having similar sensitivity characteristics with the crosstalk error [1]. Depending on the sensor-working environment, crosstalk error further affects the performance of gas detection at the system level. To improve the system performance of gas detection, both principal component analysis (PCA) and artificial neural network (ANN) are widely used in many applications [2,3] but in some cases they fail to detect gas concentration accurately [4,5].

In this paper, we propose a new method of pattern recognition and concentration measurement with gas sensor array measurement data using the independent component analysis (ICA). ICA is a high-order multivariate statistical method that seeks a linear representation of non-Gaussian data so that the output components are as independent as possible. The measurement data are obtained form a gas sensor array which consists of six different commercial MOS sensors. In simulation study, we use 5 different concentrations of monoxide (CO) and methane (CH4) as known samples. Another 5 different concentrations of the two gases are used for test samples. It has shown that the proposed method is very promising to provide more meaningful representation for gas identification.

The rest of the paper is organized as follows. Section 2 provides a brief explanation about the ICA. Experimental results are presented in section 3, and finally conclusion is given in section 4.

2. ICA model for analysis of sensor array measurement data

Fig. 1 shows a basic structure of the gas sensor array system. Each sensor j of a system produces a response x_j to input gas i. The sensor array response can be represented by a data matrix X,

$$X = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1N} \\ x_{21} & x_{22} & \cdots & x_{2N} \\ \vdots & \vdots & & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nN} \end{pmatrix} \quad (1)$$

where n is the number of gas sensors in the array and N is the sample number. In particular, the semiconductor gas sensor responds to different gases based on the chemical sensitive mechanism that produces nonzero cross terms in the matrix X. Thus, gas sensors in the system are often suffering from problems associated with nonselective property when they are used as detectors for a specific gas component [6]. As a result of the cross-sensitivity, it is difficult to identify a specific gas component accurately.

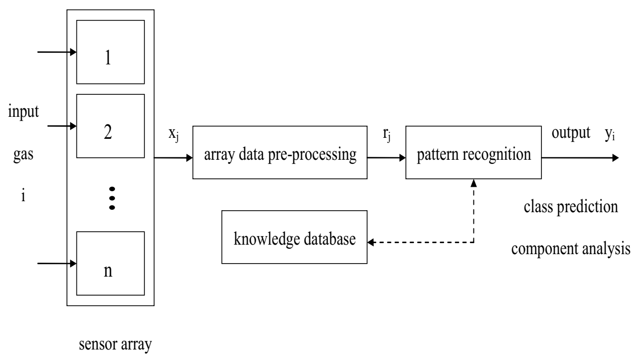

Fig. 1 Basic structure of a gas sensor array system

The ICA has drawn considerable attention to unsupervised neural learning and array signal processing as a higher-order multivariate statistical tool [7,8]. The main purpose of ICA is to recover unknown source signals form their linear mixtures so that the separated signals are statistically independent. A generic ICA model can be written as

$$x = As \quad (2)$$

where $s=[s_1, s_2, \ldots, s_m]^T$ denote m source signals, $x=[x_1, x_2, \ldots, x_n]^T$ represent n mixed signals and A is $n \times m$ mixing matrix.

Generally speaking, independent components can be derived if all the following assumptions are satisfied: source signals s are statistically independent and have non-Gaussian distributions; mixing matrix A is regular, which means that ICA exploits spatial diversity of a sensor array; the number of source signals is less than or equal to the number of sensors in a sensor array.

The problem of ICA is to set a separation matrix B which is the estimation of A^{-1}. The matrix B can be calculated by optimizing an independent criterion. Once the separation matrix B has been estimated, separated signals $y=[y_1, y_2, \ldots, y_m]^T$ that are the estimations of s can be calculated as follow.

$$y = Bx = BAs \approx s \tag{3}$$

There are many algorithms for ICA but in this study the fast fixed-point ICA (FastICA) is applied to determine the independent components. The FastICA is known as a very flexible fixed-point algorithm that performs well for almost any kind of source estimation [9].

3. Experimental Results

The gas sensor array response data correspond to mixed signals of ICA. In the simulation, we used the sensor array response data given in [10] that are obtained from a commercial semiconductor sensor array manufactured by Hanwei Electronics Ltd., Henan, China. The gas sensor array consists of 6 different tin dioxide sensors: (A) MQ-2, (B) MQ-4, (C) MQ-5, (D) MQ-6, (E) MQ-7, (F) MQ-8. For pattern recognition using the ICA model, the sensor array response of all known samples and the response of unknown gas to be predicted are processed in one procedure.

First, 5 different concentrations of CO and 5 different concentrations of CH4 as known samples are described as follows: 1) CO for concentrations 200, 600, 1000, 1400 and 1800ppm. 2) CH4 for concentrations 200, 600, 1000, 1400 and 1800ppm.

Then we use another 5 different concentrations of the two gases as test samples: 1) CO for concentrations 400, 800, 1200, 1600 and 2000 ppm. 2) CH4 for concentrations 400, 800, 1200, 1600 and 2000 ppm.

Fig. 2 illustrates the preprocessed response data of the gas sensor array having 6 sensors. The x-axis denotes the measurement numbers of different gas concentrations separated with grid lines in each plot, and the y-axis represents the sensor response to the different gas concentrations. For example, measurement numbers 1-10 in x-axis denote 10 measurements in the presence of 200ppm CO, and y-axis corresponds to the sensor response of 200ppm CO with 10 measurements. Measurement numbers 101-110 denote 10 measurements of the unknown gas to be predicted, and the ordinate denotes the sensor response to them. Particularly, measurement number 50 is the boundary dividing measurements of two different gases, i.e. measurement numbers 1-50 represent the concentration information of CO, and 51-100 denote the concentration information of CH4.

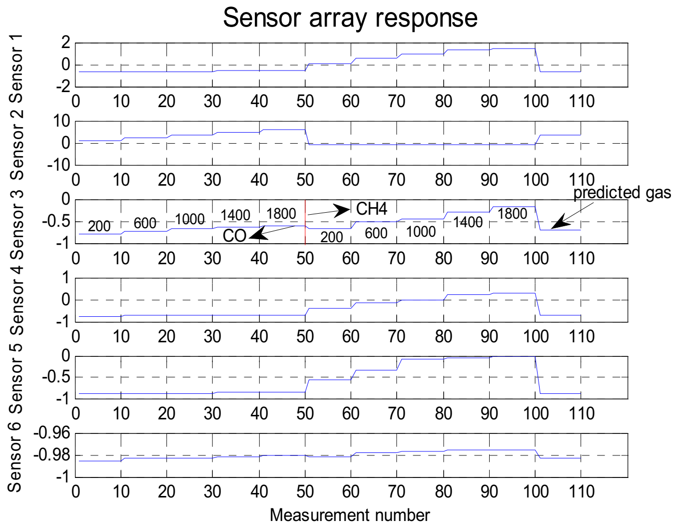

Fig. 2 Gas sensor array response data

We carried out the FastICA algorithm to get the independent components (ICs) that correspond to separated signals of ICA. Gas component analysis using this ICA-based model comprises two stages: Firstly we achieve the gas classification, and then measure the gas concentration. Here the ICA scatters plot (IC1 vs. IC2) which is an intuitional tool for gas component analysis is used. Each scatter plot displays the information of each gas component and concentration. There are 10 different CO and CH4 known sample concentrations and an unknown test gas concentration, so the number of all components in the scatter plots is 110 (10 measurements each). We only give two examples of the scatter plot for sensor array response of 800ppm CO and 1600ppm CH4, respectively. Fig. 3 shows the scatter plot of two ICs (IC1 vs. IC2) for predicted 800ppm CO. We can easily discriminate the different concentrations of CO and CH4 from the scatter plot.

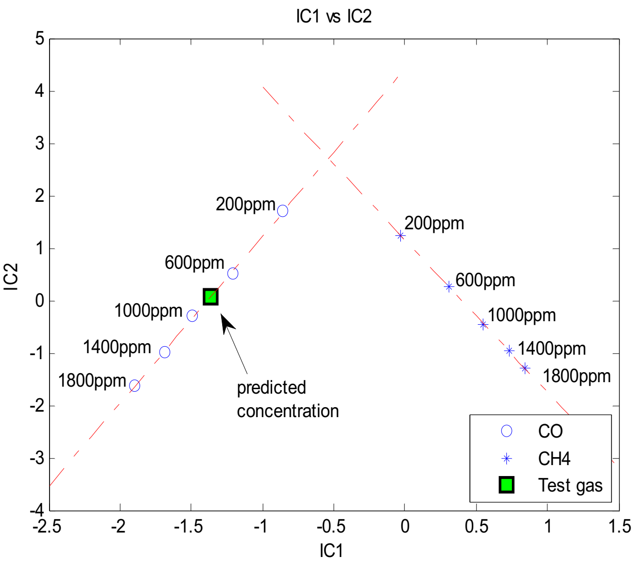

Fig. 3 ICA scatter plot with predicted 800ppm CO

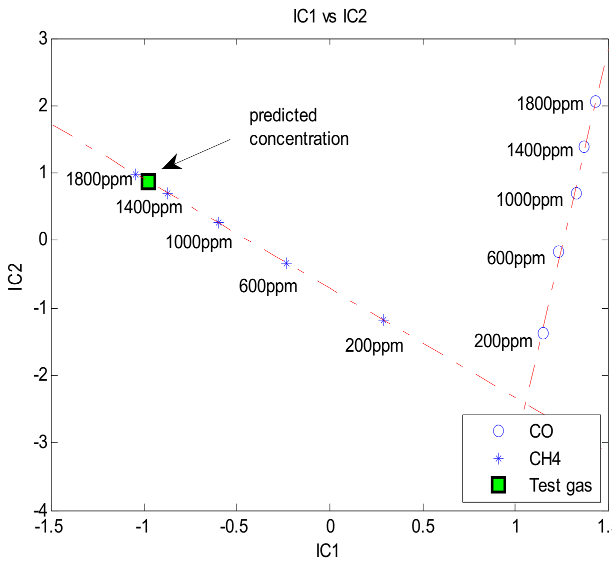

Fig. 4 ICA scatter plot with predicted 1600ppm CH4

In Fig. 3, scatters 1-50 (round dots) represent the concentration information of CO, e.g. scatters 1-10 denote CO 200ppm, 11-20 CO 600ppm, 21-30 CO 1000ppm, 31-40 CO 1400ppm and 41-50 CO 1800ppm. Similarly, scatters 60-100 (star dots) represent the concentration information of CH4, e.g. scatters 51-60 denote CH4 200ppm, 61-70 CH4 600ppm, 71-80 CH4 1000ppm, 81-90 CH4 1400ppm and 91-100 CH4 1800ppm. It is obvious that CO concentrations are almost proportionally distributed along the direction in which CO concentrations increase. In a similar way, CH4 concentrations are proportionally distributed along the direction in which CH4 concentrations increase. Scatters 101-110 represent the concentration information of the predicted gas. According to the scatter position of the predicted gas in the ICA scatter plot, we may judge the unknown gas class. For example, since scatters 101-110 (square dots) lie in the direction of CO concentration distribution, the test gas is CO, otherwise it will be CH4.

Fig. 4 illustrates the ICA scatter plot of the predicted 1600ppm CH4. We can easily obtain the information of gas class and concentration. In fact, the ICA scatter plots are usually not the same in each computing process. It is because of the ICA indeterminacy which changes the scale and order of estimated ICs. However, although the relative positions of scatters are different in different scatter plots, the space distances of scatters are not changed so each result of gas recognition and concentration measurement is always same.

From figures 3 and 4, we can see that the FastICA-based method achieves the qualitative and quantitative gas component analysis according to the information of scatter, so it is very promising for gas identification with estimation of its concentration.

4. Conclusion

Commercial metal oxide semiconductor gas sensor array system is generally subject to significant cross-sensitivity due to poor selectivity for different gases. This crosstalk greatly affects the accuracy of gas recognition and concentration measurement. In this paper, we explore the ICA as a novel modeling technique to process the gas sensor array response data. The processing consists of two stages: the first stage recognizes the gas class; the second one measures the gas concentration using the scatter plot of independent components. Simulation results demonstrate that the ICA-based method show good performance on the qualitative and quantitative gas component analysis.

References

[1] E.L.Hines, E.Llobet and J.W.Gardner, "Electronic Noses: A Review of Signal Processing Techniques," *IEE. Proc. Circuits Devices Syst.*, vol. 146, no. 6, pp. 297-310, 1999.

[2] J. W. Gardner, "Detection of Vapours and Odours from A Multisensor Array Using Pattern Recognition Part 1. Principle Component and Cluster Analysis," *Sensors and Actuators B*, vol. 4, pp. 109-115, 1991.

[3] M. Penza and G. Cassano, "Application of Principal Component Analysis and Artificial Neural Networks to Recognize the Individual VOCs of Methanol/2-propanol in Binary Mixture by SAW Multi-sensor Array," *Sensors and Actuators B*, vol. 89, pp. 269-284, 2003.

[4] M. Bicego, G. Tessari, G. Tecchiolli and M. Bettinelli, "A Comparative Analysis of Basic Pattern Recognition Techniques for the Development of Small Size Electronic

Nose," *Sensors and Actuators B*, vol. 85, pp. 137-144, 2002.

[5] R. E. Shaffer, S. L. Rose-Pehrsson, R. A. Mcgill, "A Comparison Study of Chemical Sensor Array Pattern Recognition Algorithms," *Analytica Chimica Acta*, vol. 384, pp. 305-317, 1999.

[6]A. Bermak and S. B. Belhouari, "Bayesian Learning Using Gaussian Process for Gas Identification," *IEEE Transactions on Instrument and Measurement*, vol. 55, no. 3, pp. 787-792, 2006.

[7]A. Cichocki and S. I. Amari, *Adaptive Blind Signal and Image Processing: Learning Algorithms and Applications.* New York: Wiley, 2002.

[8]A. Paraschiv-Ionescu, C. Jutten and G. Bouvier, "Source Separation Based Processing for Integrated Hall Sensor Arrays," *IEEE Sensors Journal*, vol. 2, no. 6, pp. 663-673, 2002.

[9]A. Hyvarinen, J. Karhunen and E.Oja, *Independent Component Analysis.* New York: Wiley, 2001.

[10]C. Y. Shi, "Study and Design on Multiple Gases Analysis System based on Artificial Neural Network," *Master Thesis*, Ji Lin University, May 2004 (in Chinese).