

Pattern Recognition Applied to Analysis of Gas Sensors' Array Data

P. MARCZYŃSKI*, A. SZPAKOWSKI, C. TYSZKIEWICZ AND T. PUSTELNY

Department of Optoelectronics, Silesian University of Technology

B. Krzywoustego 2, 44-100 Gliwice, Poland

For an array of eight chemoresistive gas sensors a computer pattern recognition system was built. Multivariate data analysis was performed for measurements of three gases' dilutions: hydrogen (H_2), methane (CH_4), and carbon monoxide (CO). The pattern recognition system included a feature subset selection algorithm involving PCA and objective function. Dimensionality reduction was applied to two kinds of patterns: three aforementioned gases and six different concentrations of hydrogen. For patterns of the three gases, classification tests were performed using k -NN algorithm and N -fold based validation method.

PACS: 02.50.Sk, 07.07.Df

1. Introduction

Pattern recognition is commonly applied to machine olfaction domain as an important component of device called an electronic nose [1–7]. The measurements typically come from an array of many gas sensors of miscellaneous selectivity. The point is to gather superfluous data about examined quantity and then extract distinctive information discriminating observed objects. However, before the particular measurement, a pattern recognition system needs to be trained with known samples, as recognition stands for “an awareness that something perceived has been perceived before”**. Patterns then would be the totality of measurements for known objects, which the system has been trained with and is meant to recognize.

In context of gas sensors array the pattern recognition mechanism can be divided into four steps: preprocessing, feature extraction, dimensionality reduction, and classification [2]. First step consists of raw measurements' noise filtering, baseline compensation (sensors' drift), scaling and normalization [2]. Next — from sensors' transient responses — features are extracted, e.g. maximum response [3], on integral [3], value after some time lapse [3], values of derivatives [3, 8] etc. Obtained this way feature vectors are input data for reduction of dimensionality, that is (at this stage) in order of hundreds (number of sensors multiplied by number of features). Dimensionality reduction is composed of feature projection (PCA is widely applied [1, 3, 4, 6–9]) and feature subset selection mechanisms [1, 3, 10–13] and aims to find features that carry the most patterns discriminative information. The

last step is classification — creation of decision boundaries in reduced feature space — in order to make the assignment of an observation to particular pattern class possible.

2. Experimental setup

The elements of sensors array were semiconductor chemoresistive Figaro TGS gas sensors based on tin dioxide (SnO_2). The array was composed of eight sensors, which are specified in Table I. Operating temperature of measuring chamber, where sensors were installed, was maintained on constant level.

Each measuring series was a series of transient response cycles for different gas concentrations obtained with gas and synthetic air flow control in the chamber. Depending on measuring series the flows increased in every following cycle or changed alternately. A full measuring cycle was divided in two intervals: adsorption and desorption of gas, during which the sensors transient responses attained constant level.

The measured quantity was conductivity, which acquired an automated data acquisition system with 10 Hz sample rate, logging start/end time points of adsorption and desorption sub-cycles.

3. Experimental results

The measurements were performed for dilutions of hydrogen (H_2), methane (CH_4), and carbon monoxide (CO). Sensors' transient responses were measured for six different concentrations (flows) in synthetic air: 10, 20, 35, 50, 65, 80 ml/min, at total flow 100 ml/min.

In the data preprocessing step the signal was filtered with the Hanning windowing function [13], next the sensors' baseline was compensated. Two approaches of baseline compensation were applied: differential and fractional [2], doubling thereby the number of afterwards extracted features.

* corresponding author; e-mail: Przemek.Marczynski@gmail.com

** definition of *recognition* from The Free Dictionary by Farlex Inc.

TABLE I
List of chemoresistive Figaro gas sensors installed in the gas sensor array. Selectivity information in compliance with data sheets provided by Figaro Inc.

Sensor	Destination	Sensor	Destination
TGS 813	C ₃ H ₈ , C ₄ H ₁₀ , LP gas	TGS 2602	air contaminants
TGS 842	CH ₄ , natural gas	TGS 2620	organic solvents, alcohol vapors
TGS 800	air contaminants	TGS 2610	C ₃ H ₈ , C ₄ H ₁₀ , LP gas
TGS 2611	CH ₄ , natural gas	TGS 2600	air contaminants

16 generic features were designed to be extracted from each measuring cycle, that multiplied by 8 sensors in the array and two methods of baseline compensation resulted with 256 dimensional feature vector on the input of dimensionality reduction stage.

Within acquired experimental data two groups of patterns were examined separately. First the 3 gases (H₂, CH₄, and CO) were treated as patterns. The second patterns group were 6 concentrations of H₂.

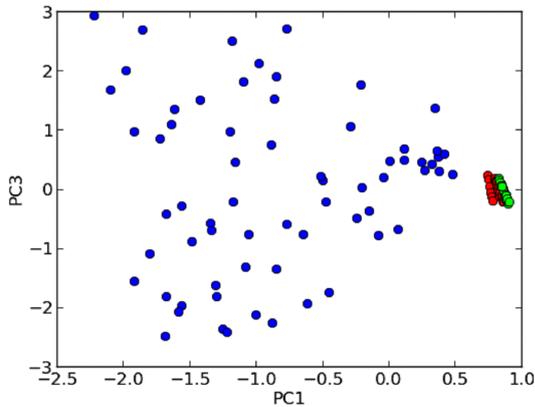


Fig. 1. Full feature set (256 features). Poor separation of 3 gases' patterns.

After feature extraction the dimensionality reduction of feature space was performed. The plus-L minus-R (LRS) [10, 13] algorithm was applied with an objective function as quality criterion — a filter approach [12] — based on a minimization of within-cluster variances. Before each feature set quality evaluation in intermediate steps of feature subset selection, a projection onto 3-dimensional space with PCA algorithm was performed. As a result of dimensionality reduction of feature vector through selection, an optimal feature subset was obtained. Projection of the feature vector by means of PCA allowed separation of pattern clusters.

In case of 3 gases' patterns the number of features picked by selection mechanism was 85. The optimal subset separating clusters of 6 hydrogen concentrations counted 80 features. Visualizations of dimensionality reduction results before and after feature subset selection are presented in Figs. 1, 2, 3 and 4. As the figures show, due to applied dimensionality reduction methods, separation and aggregation of pattern clusters on scatter plots

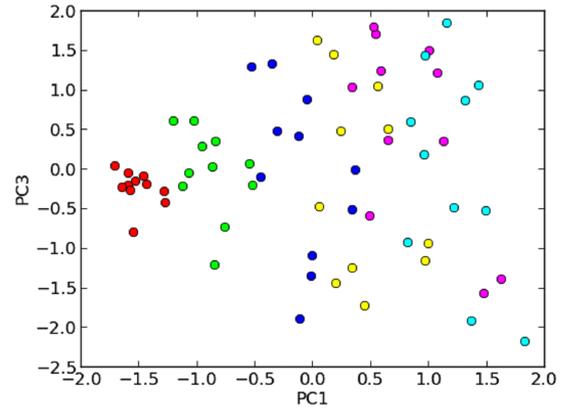


Fig. 2. Full feature set (256 features). Poor separation of 6 hydrogen concentrations' patterns.

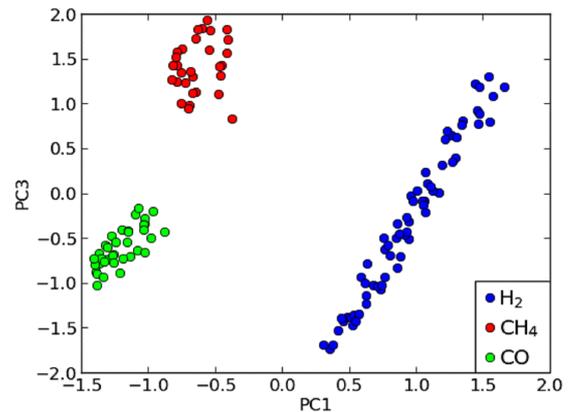


Fig. 3. Optimal feature set for 3 gases' patterns (85 features). Clusters aggregation significantly increased.

gained significant improvement, enabling further calculation of decision boundaries.

For all above data visualizations feature space was projected by means of PCA onto 3D space. 2D figures show relations between two principal components: $PC3 = f(PC1)$.

Finally, validation of the pattern recognition system was performed for the patterns of 3 gases. N -fold validation [2] based method was applied. All measuring series except for the one, to which the current test sample belonged, were assumed as the training dataset. The ap-

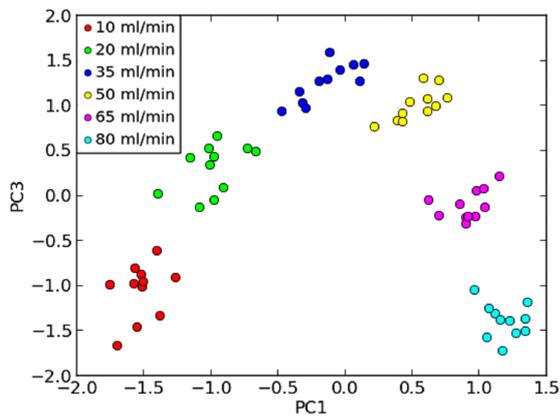


Fig. 4. Optimal feature set for 6 hydrogen concentrations' patterns (80 features). Clusters aggregation significantly increased.

plied classifier was k -nearest neighbors algorithm [2] (for $k = 1, 3, 5$). Table II presents classification results as ratio of correctly classified samples to total number of samples in particular tested measuring series.

TABLE II

Classification results for patterns of 3 gases by means of k -NN algorithm.

Meas. series	Gas	Samples	Recognition ratio (correctly classified samples)		
			1-NN	3-NN	5-NN
#1	H ₂	18	100%	100%	100%
#2	CH ₄	6	100%	100%	100%
#3	H ₂	12	100%	100%	100%
#4	H ₂	12	100%	100%	100%
#5	H ₂	24	100%	100%	100%
#6	CH ₄	24	37.5%	58.3%	58.3%
#7	CO	12	100%	100%	100%
#8	CO	12	100%	100%	100%
#9	CO	12	100%	100%	100%

4. Conclusions

The results of research show that the developed dimensionality reduction mechanism is effective for gases discrimination as well as their concentrations. As it was presented in case of hydrogen concentrations, the system allows separation up to six pattern classes.

As it can be seen in Table II, recognition ratio for measuring series #6 is significantly lower comparing to others. In this particular case the only training dataset for test gas (CH₄) was series #2 counting only 6 samples. It is very low for a training data set, and hence had influence on pattern clusters aggregation and positioning in feature space at dimensionality reduction stage.

It is worth to notice that a drawback of k -NN based classifier is weak generalization, so it is presumed that

obtained results can be overfitted. In future, it may lead to misclassification of tested objects. For that reason it is planned to apply an artificial neural network based classifier.

The applied objective function yields accurate results, however doesn't guarantee good quality evaluation of any data. To solve this problem, a wrapper approach [12] of objective function together with k -NN voting rule in intermediate selection steps [1, 3] will be a part of further research. The selection algorithm is also planned to be improved in order to result in less numerous optimal feature subset. It is considered to implement dynamic parameters L and R of LRS method (a floating selection). Alternatively, randomized selection methods will be considered.

Moreover, in further perspective, it is planned to adapt the pattern recognition system to analyze experimental data from source different than the gas sensor array.

Acknowledgments

The researches were financed in the frame of the project No. OR 00017912 of the State Committee for Scientific Researches and Developments (NCBiR).

References

- [1] M. Pardo, G. Sberveglieri, *Sensors Actuators B* **123**, 437 (2007).
- [2] R. Gutierrez-Osuna, T. Nagle, *IEEE Trans. Systems, Man Cybernetics — Part B: Cybernetics* **29**, 626 (1999).
- [3] T. Sundic, S. Marco, A. Perera, A. Pardo, J. Samitier, P. Wide, in: *5th Seminar on Neural Network Applications in Electrical Engineering, NEUREL, Univ. Troy, New York*, 2000, p. 201.
- [4] C. Tyszkiewicz, T. Pustelny, *Opt. Appl.* **34**, 507 (2004).
- [5] T. Pustelny, J. Ignac-Nowicka, B. Jarzabek, A. Burian, *Opt. Appl.* **34**, 561 (2004).
- [6] Y. Yin, X. Tian, *Sensors Actuators B* **124**, 393 (2007).
- [7] P.U. Kurup, *2008 IEEE Conf. Technologies for Homeland Security*, Massachusetts Univ., Lowell (MA) 2008, p. 144.
- [8] S. Roussel, G. Forsberg, V. Steinmetz, P. Grenier, V. Bellon-Maurel, *J. Food Eng.* **37**, 207 (1998).
- [9] A.H. Gómez, J. Wang, G. Hu, A.G. Pereira, *Sensors Actuators B* **113**, 347 (2006).
- [10] K.Z. Mao, *IEEE Trans. Systems, Man Cybernetics — Part B: Cybernetics* **34**, 629 (2004).
- [11] T. Pustelny, A. Opilski, B. Pustelny, *Acta Phys. Pol. A* **114**, A-183 (2008).
- [12] E. Cantú-Paz, S. Newsam, C. Kamath, in: *Int. Conf. on Knowledge Discovery and Data Mining*, Seattle, WA, USA, 2004, p. 411.
- [13] A. Szpakowski, C. Tyszkiewicz, T. Pustelny, *Acta Phys. Pol. A* **114**, A-239 (2008).