Available online at www.sciencedirect.com
SCIENCE DIRECT®
Talanta 64 (2004) 609–617
Talanta
www.elsevier.com/locate/talanta

Recognition of benzene, toluene and xylene using TGS array integrated with linear and non-linear classifier

Andrzej Szczurek, Monika Maciejewska*

Ecologistics and Atmosphere Protection Group, Institute of Environmental Protection Engineering, Wroclaw University of Technology, Pl. Grunwaldzki 9, 50-377 Wroclaw, Poland

Received 5 December 2003; received in revised form 4 March 2004; accepted 22 March 2004
Available online 19 May 2004

Abstract

Three volatile organic compounds (VOCs): benzene, toluene and xylene were measured with an array of six Taguchi gas sensors in the air with variable humidity content. The recognition of single compounds was performed, based on measurement results. The principal component analysis (PCA) pointed at humidity as the main classification factor in the measurement data set. The linear discriminant analysis (LDA) was applied to overcome this drawback and enforce classification with respect to benzene, toluene or xylene. It was shown that discriminant function analysis (DFA), which is an LDA method allowed for 100% success rate in test samples recognition of benzene. It did not allow for accurate recognition of test samples of toluene or xylene. Following, the non-linear classifier, radial basis function neural network (RBFNN) was applied. A specific configuration of input ‘s was found, which provided for successful recognition of each single compound: benzene, toluene or xylene in air with variable humidity content.

Keywords: Sensor array; DFA; RBF; VOC

1. Introduction

The pollution of atmosphere as a result of human activities is recently one of the most important problems all over the world. This fact stimulates the development of measurement methods and techniques for air pollutants monitoring. Among various analytical devices, gas sensors are particularly useful, first of all in the case of in situ, on line or remote measurements. The gas sensors may operate on the base of different physical and chemical methods. For example, in a great number of detectors for toxic and inflammable gases semiconducting properties of metal oxides are explored. These devices are relatively small, inexpensive (compared with optical or electrochemical gas sensors), convenient in use and sensitive. They also have disadvantages, such as considerably large power consumption, lack of reproducibility, long-term stability (drift), sensitivity to poisoning and ambient conditions, and insufficient selectivity. In order to improve reliability of semiconducting gas sensors and broadening their applications different methods are proposed. They are based on: (1) doping (the role of dopants and additives is to cause the increase of surface adsorption or to improve electrical and mechanical properties of gas sensor) [1] (2) filters (they reduce the sensitivity to interfering gases) [2] (3) catalysts (catalytic metals and oxides lead to the decrease in the operation temperature and to the enhancement of sensitivity and selectivity to different gases) [3] (4) surface modification [4] (5) different operating temperature of the sensor [5] (6) dynamic mode of operation (“time dependent” response of the sensor, the measuring signal is modulated by different factors) [6] (7) multisensor arrays, which combine sensor elements having different sensitivity to each component of the mixture [7], and (8) pattern recognition techniques [8–12].

In this paper, we investigate the classification problem of three single volatile organic compounds (VOCs): benzene, toluene and xylene in air containing one of these compounds and various amounts of water vapour, based on sensor array measurements. These compounds were chosen because they belong to the group of most dangerous air pollutants. Their recognition is very important due to differences in toxicity of these compounds and their impact on human health and the environment. Selected feature extraction and pattern recog-

* Corresponding author. Tel.: +48-713202500; fax: +48-713203599.
E-mail address: monika.maciejewska@pwr.wroc.pl (M. Maciejewska).

0039-9140/$ – see front matter © 2004 Elsevier B.V. All rights reserved.
doi:10.1016/j.talanta.2004.03.036

nition techniques were employed to solve the classification problem. We used principal component analysis (PCA), linear discriminant function (LDF) and radial basis function neural network (RBFNN).

2. Experimental

In our study, sensor array was used for the recognition of benzene, toluene and xylene in air with variable humidity content. It consisted of six Taguchi gas sensors from Figaro Engineering, Japan, which were chosen because of their large commercialization, relatively low price and high sensitivity. The following TGS were applied in this work: TGS800, TGS822, TGS824, TGS825, TGS880, and TGS883. The experimental set-up was shown in Fig. 1.

The sensor array was mounted in a flow type gas chamber equipped with two ports for the gases (inlet and outlet) and a head with wires. The function of sensor array was responding to gases and transformation of chemical information into the electrical signal. The sensor array was connected to voltage supplier and measuring device. Both were regarded as separate elements of the measurement system.

The measuring device consisted of reference resistors, digital multimeter with data acquisition card and PC. We used the potential divider circuit, each connected to 12 V circuit supply. The sensors responses were measured in the form of the voltage of the reference resistors. A digital multimeter with data acquisition card was used to measure and transfer the change of output voltage to PC. TGS were operated in working temperature of approximately 350 °C, by applying constant 5 V supply to heaters.

The chamber was connected by a teflon gas line with the installation for the preparation of zero air (purified ambient air) and gas mixtures (samples). Zero air came from the generator made by Horiba. This device consisted of the compressor and cartridges filled with silica gel, activated carbon, soda lime and molecular sieves. Ambient air passing through these cartridges was purified. Water vapor was not completely removed from the zero air by silica gel, but its influence on electrical conductivity of semiconductor was negligible. The zero air generator was designed for the dynamic and continuous supply of pure air. The gas flow rate was controlled by a mass flow controller.

Gas mixtures were prepared by an evaporation method. A desired amount of water and a volatile organic compound was injected as a liquid into a glass coil and then vaporized in a stream of pure air. This gas mixture was collected in a Tedlar bag. The flow rate of air was precisely adjusted. Therefore, the concentration of volatile organic compound in the bag was known. Finally, the gas mixture from the Tedlar container was pumped into the chamber and the gas flew over the sensor array at constant flow rate 2 $dm^3\,min^{-1}$. Each sensor response was measured after four minutes from the start of mixture flow. Between exposures, the sensor array was kept in a stream of pure air.

Sensor array responses were measured for benzene, toluene and xylene in air. The concentration of these

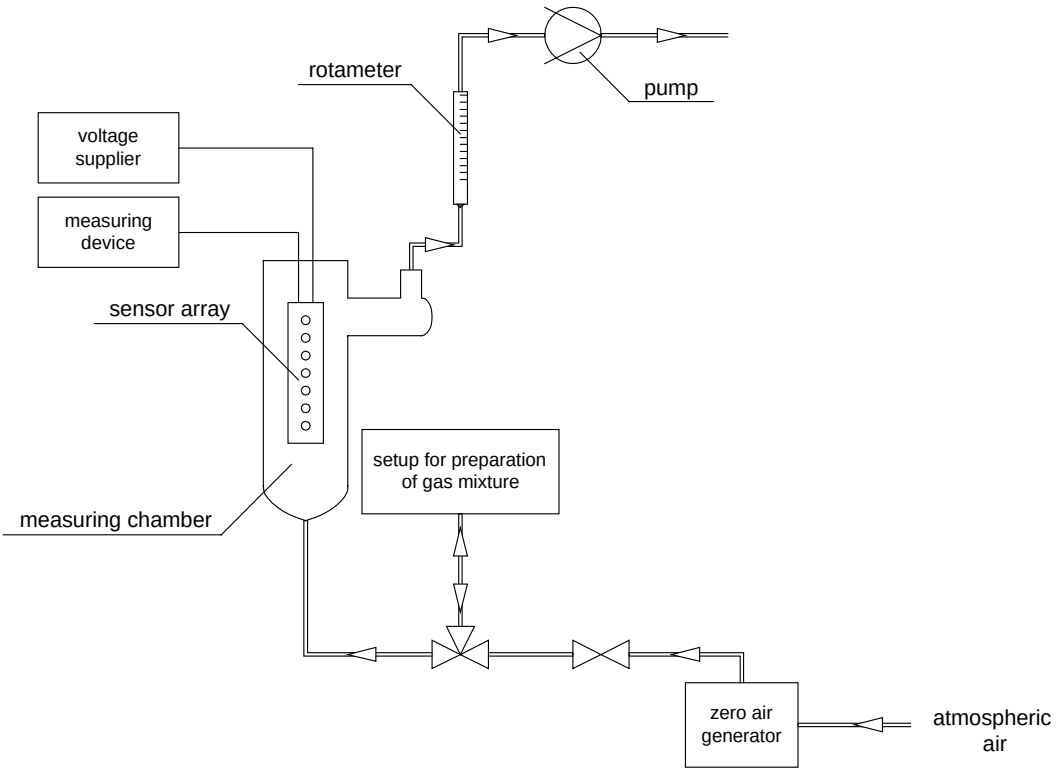

Fig. 1. The structure of the experimental set-up.

compounds was changed from 3 to 73 ppm. The air humidity content was also changed. We performed measurements at four air humidity levels (dry air, 4 g(H_2O) m^{-3}, 10 g(H_2O) m^{-3}, water vapor saturated air).

3. Results and discussion

3.1. Feature extraction with PCA

The experimental data was examined with PCA in order to visualize response patterns in the feature space of principal components. PCA is a coordinate transformation, which reduces the redundancy within the data by creating a new series of components. The resulting components are often more interpretable than the original ones. Usually, the first two or three principal components account for most of the variance in the data set. Therefore, PCA may be used as dimensionality reduction technique. To begin the transformation, the covariance matrix $\boldsymbol{C}$ or the correlation matrix of the original data is found. Using the covariance matrix, the eigenvalues λ_i are obtained from

$$|\boldsymbol{C} - \lambda_i \boldsymbol{I}| = 0 \tag{1}$$

where $i = 1, 2, \ldots, n$ and n is the total number of original components, and $\boldsymbol{I}$ is an identity matrix. The eigenvalues are equal to the variance of each corresponding component. The eigenvectors e_i define the axes of the components and are obtained from

$$(\boldsymbol{C} - \lambda_i \boldsymbol{I}) e_i = 0 \tag{2}$$

The principal components are then given as

$$\mathrm{PC} = \boldsymbol{TD} \tag{3}$$

where $\boldsymbol{D}$ is the digital number matrix of the original data, and $\boldsymbol{T}$ is the transformation matrix given by

$$\boldsymbol{T} = \begin{bmatrix} e_{11} & \cdots & e_{1n} \\ \vdots & \ddots & \vdots \\ e_{n1} & \cdots & e_{nn} \end{bmatrix} \tag{4}$$

PCA generated components are uncorrelated and ordered by decreasing variance. The correlation matrix of the transformed data is a diagonal matrix. Its elements are comprised of eigenvalues. The transformed data points are linear combinations of original data values weighted by eigenvectors. Each column of the eigenvector matrix represents the weights applied to the data point of the corresponding data vector to build each respective principal component. The percentage of the total variance in each of the components is given by

$$\text{variance}_i\ (\%) = \frac{\lambda_i 100}{\sum_{k=1}^{n} \lambda_k} \tag{5}$$

The first component has the maximum signal-to-noise ratio and the largest percentage of the total variance. Each subsequent component contains the maximum variance for any axes orthogonal to the previous component.

PCA is commonly applied as an "unsupervised" linear technique of feature extraction, which analyses the structure inherent in the data. The plot of response patterns in the space of major principal components is frequently considered as the first indicator of success or failure in data classification task [8,10,11,13].

The PCA algorithm employed in the study used *princomp* function implemented in MATLAB [15].

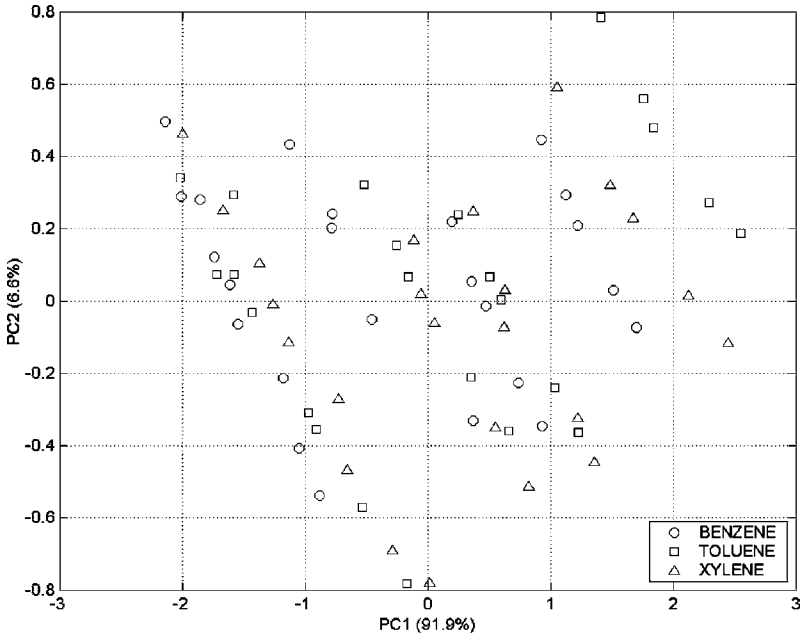

Fig. 2. PCA of 72 patterns from the experimental data set; all sensors used for calculation of principal components; marking for VOCs.

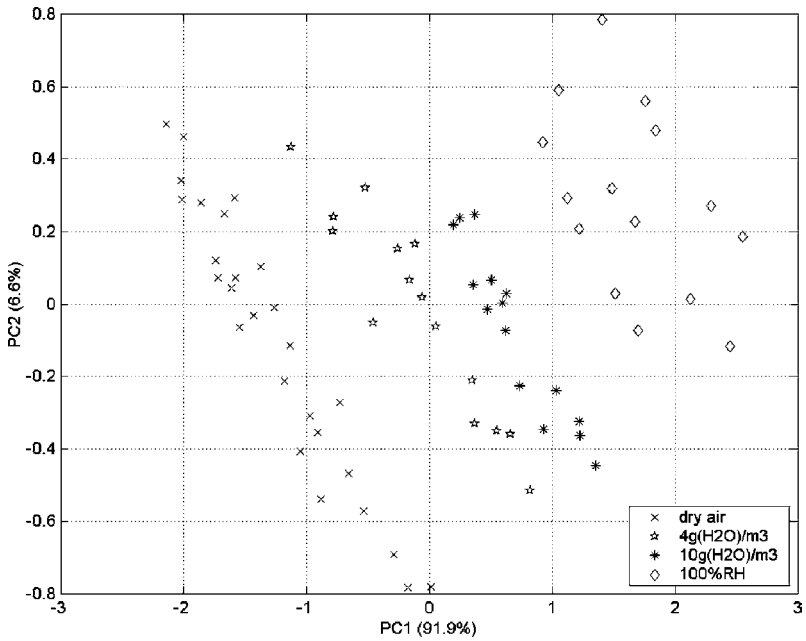

Fig. 3. PCA of 72 patterns from the experimental data set; all sensors used for calculation of principal components; marking for air humidity level.

The basic measurement data set consisted of 72 pattern vectors collected using sensor array (TGS800, TGS822, TGS824, TGS825, TGS880, and TGS883). It consisted of three groups of pattern vectors, 24 elements per group. Each group contained measurement results obtained for one single compound in air, in full concentration range, at four humidity levels. PCA was employed on six dimensional pattern vectors (all sensors considered) and on five dimensional pattern vectors (humidity sensor TGS883 excluded). These two cases were analyzed, as our previous work [16] indicated that including humidity sensor responses in experimental data set influenced the VOC classification performance. Specifically, it was shown to improve alcohols classification rate. Prior to the analysis pattern vectors were centered and normalized. Figs. 2 and 3 visualize pattern space for complete set of sensors, using first two principal components. Different marking was adopted in these two figures although they visualize exactly the same data. The marking used in Fig. 2 refers to three VOCs: benzene, toluene and xylene, while the marking used in Fig. 3

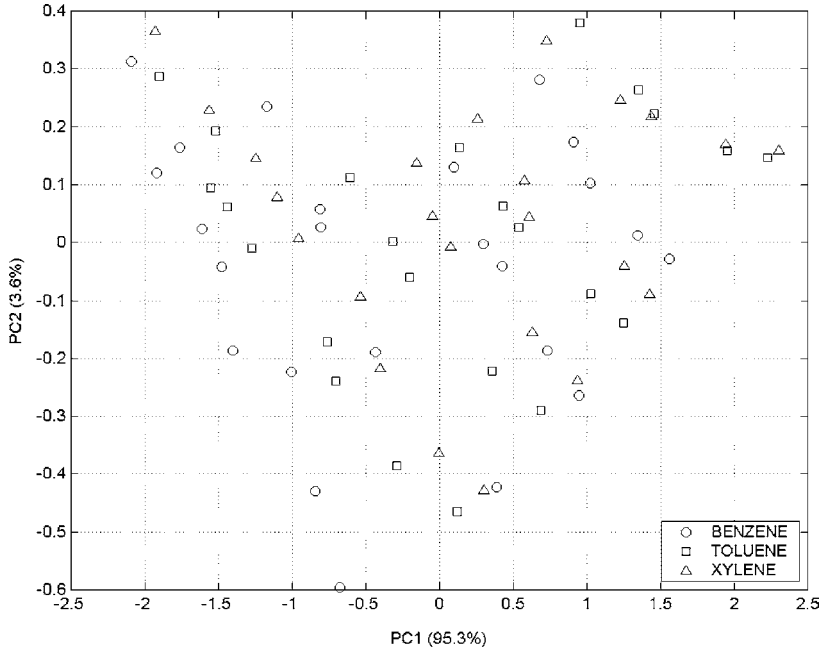

Fig. 4. PCA of 72 patterns from the experimental data set; humidity sensor excluded prior to calculation of principal components, marking for VOCs.

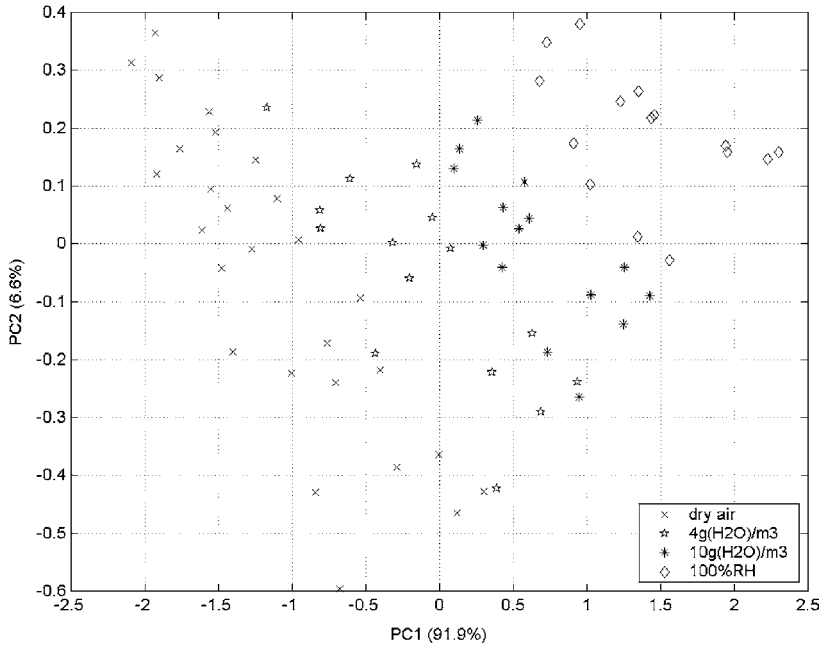

Fig. 5. PCA of 72 patterns from the experimental data set; humidity sensor excluded prior to calculation of principal components; marking for air humidity level.

refers to four air humidity levels (dry air, $4\,g(H_2O)\,m^{-3}$, $10\,g(H_2O)\,m^{-3}$, water vapor saturated air).

As could be seen in Fig. 2, four distinct groups of points were formed, but clearly these did not indicate VOCs of interest. From the comparison of Figs. 2 and 3, it appeared that four data point groups represented four levels of water vapor content in air. Apparently, the predominant fraction of variance in experimental data set (PC1 91.6%) was associated with the humidity content in the measured gas. The inherent data structure pointed at the water vapor, not VOCs as the main discrimination factor for our data set.

Based on Figs. 4 and 5, removing humidity sensor responses from data set did not induce positive regrouping in feature space. Distinct groups of points for benzene, toluene and xylene were not formed. It could only be noticed that the main factor (PC1 95.35%) less clearly indicated gas mixture humidity change, as different humidity groupings overlapped. Nevertheless, the water vapor related information was strongly represented in measurements of other gas sensors.

3.2. Linear discriminant analysis with DFA

Considering PCA results, the discriminant analysis was employed by means of discriminant function analysis (DFA). It is a commonly used technique for data classification and dimensionality reduction. The prime difference between LDA and PCA is that PCA does more of feature classification and LDA does data classification. In PCA, the shape and location of the original data sets changes when transformed to a different space whereas LDA doesn't change the location but only tries to provide more class separability and draw a decision region between given classes.

Our aim was a good discrimination of classes, so the class-dependent transformation was applied for data transformation and classification. This approach maximizes the ratio of between-class variance to the within-class variance in any particular data set thereby guaranteeing maximal separability. A set of feature vectors a_i is searched for, that maximize the following criterion

$$J = \frac{a^{\mathrm{T}} S_{\mathrm{B}} a}{a^{\mathrm{T}} S_{\mathrm{w}} a} \tag{6}$$

where $i = 1, \ldots, L$ and L is the number of classes. The normalization constraint is imposed on the criterion evaluation, that class-centralized vectors in the transformed space are uncorrelated. The within class scatter S_{W} is given by

$$S_{\mathrm{W}} = \sum_{i=1}^{L} \frac{n_i}{n} C_i \tag{7}$$

where $i = 1, \ldots, L$ and C_{i} are covariance matrices of each class (with n_i samples). The in between class scatter is given by

$$S_{\mathrm{B}} = \sum_{i=1}^{n} \frac{n_i}{n} (m_i - m)(m_i - m)^T \tag{8}$$

where m_i is the class mean and m the whole data set mean.

The solution obtained by maximizing criterion Eq. (6), defines the axes of the transformed space. The rank of S_{B} is at most $L - 1$, therefore the projection will be onto a space of at most $L - 1$ dimensions. Once the transformations are completed using the LDA transforms, Euclidean distance or RMS distance is used to classify data points.

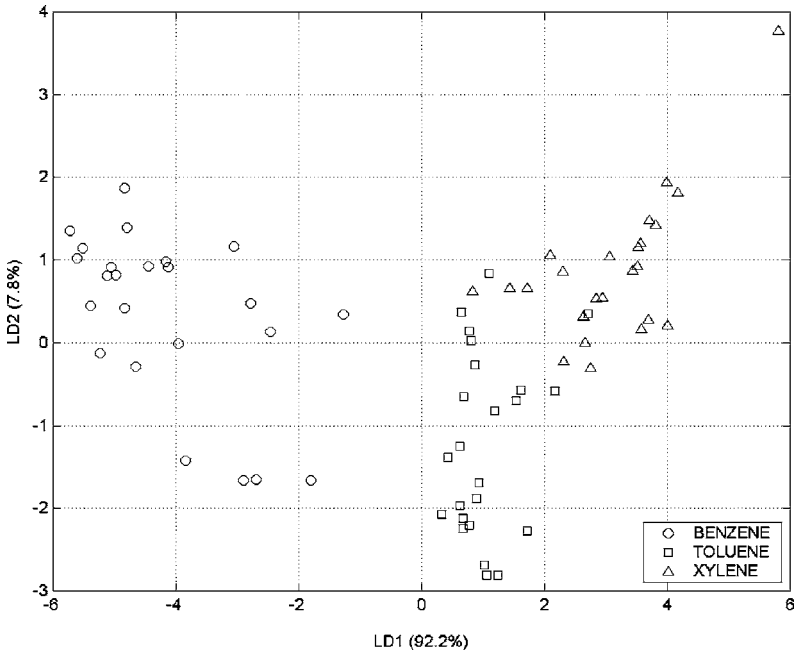

Fig. 6. Scores plot for LDA classification of 48 element test set; all sensors used to generate discriminant functions.

Thus, for L classes, L Euclidean distances are obtained for each test point. The smallest Euclidean distance among the L distances classifies the test vector as belonging to the adequate class [7,9,10,14].

In our analysis a grouping factor was assigned to each data vector in measurement data set, depending on VOC it represented. Three grouping factors were used: “a” for benzene, “b” for xyelne, and “c” for toluene.

The DFA algorithm employed in the study was the *lda* function implemented in *R* [17].

The DFA was applied to the same data set, for which PCA was performed. Fig. 6 visualizes separation of pattern vectors, using discriminant functions for complete set of sensors. The graph in Fig. 7 shows classification results, when the humidity sensor measurements were removed from the data set. Both cases were considered to see if the enforced classification scheme would suffer from the humidity oriented classification profile of the data, strongly exhibited by humidity sensor responses.

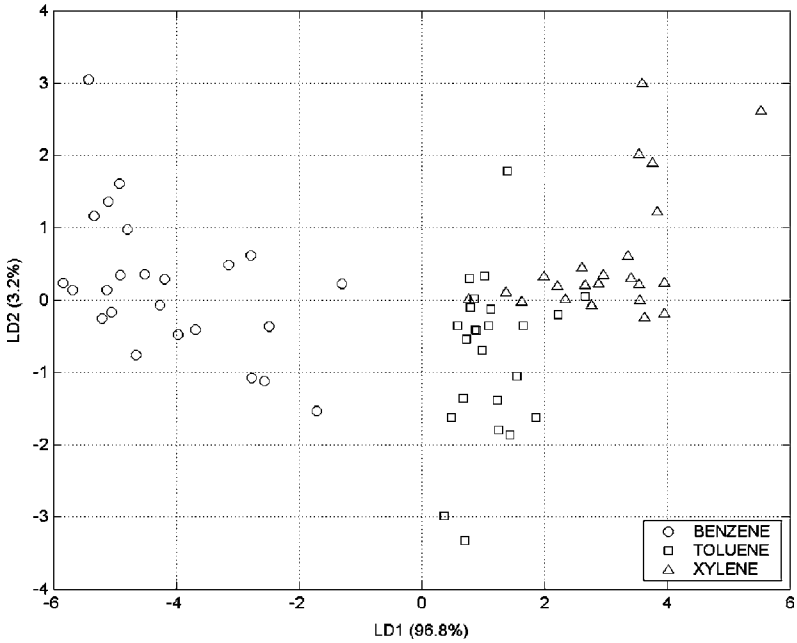

Fig. 7. Scores plot for LDA classification of 48 element test set; humidity sensor excluded prior to calculation of discriminant functions.

It was apparent, that in both graphs (Figs. 6 and 7) data points segregated into two distinct, non-overlapping groups: one for benzene, one for xylene and toluene together. Also toluene associated and xylene associated groupings were formed, but these partially overlapped. It held true irrespective of including measurements from the humidity sensor in the data set. Based on results obtained it was anticipated that benzene could be successfully recognized from toluene and xylene together, using DFA.

The modified leave-one-out cross validation was applied for evaluation of DFA discriminatory power. Classical *n*-fold cross validation, for the data set consisting of *n* data vectors, implies that the model is trained *n* separate times on all data, except for one data vector and a prediction is done for that vector. The left out vector is changed in each run. The classification errors are accumulated to give the average error across *n* trials.

Our modification was proposed due to understanding that DFA has very limited extrapolation attributes and using extreme data vectors for model evaluation in leave-one-out procedure may unjustifiably suppress the assessment of classifier performance. It was also recognized that extreme data vectors should in principle be included in the training data set, as these determine the space where the developed model is valid. Therefore in our version of cross validation extreme data vectors were excluded from testing in leave-one-out procedure, but these were left in the training set.

The full data set consisted of 72 elements. In a subset of 24 data vectors associated with one compound eight were extreme vectors, corresponding to minimum and maximum compound concentrations at four humidity levels. These eight extreme vectors for each compound i.e. 24 vectors in total, were excluded from leave-one-out procedure. Remaining 16 vectors for each compound i.e. 48 vectors in total were tested in leave-one-out procedure. Therefore our procedure modification resulted in following: In every training run 71 learning vectors (out of 72) were used, but there were only 48 runs, as there were 48 vectors to be checked if classified properly. The DFA classification results for benzene, toluene and xylene are summarized in Table 1.

The performance columns indicated the number of correctly classified patterns out of 16 test patterns for each VOC. Apparently benzene was always discriminated correctly. Most samples of toluene (15/16) and xylene (15, 14/16) were also classified well. The capability of sensor array to distinguish among VOCs was recognized, in spite of the fact that major variability in the data set was related to the air humidity level, as shown by PCA. From the point of view of practical need to recognize toxic benzene (ambient air concentration standard assigned in Poland [18]) from significantly less toxic toluene or xylene (ambient air concentration standard not assigned in Poland [18]), the DFA classification performance was satisfying. Nevertheless, our linear discriminator occasionally gave faulty classification of test samples for the two target gases: toluene and xylene.

Table 1
Classification performance of DFA in testing mode for complete set of sensors and without considering humidity sensor

VOC	Performance (complete set of sensors)	Performance (without humidity sensor)
Benzene	16/16	16/16
Toluene	15/16	15/16
Xylene	15/16	14/16

3.3. Non-linear discriminant analysis—RBF

The next step of advancement in classification methods to linear discriminators are non linear classifiers. The non linear discriminant analysis is willingly performed with radial basis functions (RBF), which is simple and fast to implement algorithm. In supervised classification applications RBF is used to construct nonlinear discrimination function for each class. In theory RBF is an absolute classifier i.e. it may generate as many classes as the number of input vectors. On the other hand, this exactness may lead to poor generalization in pattern recognition problems [11,14,19].

RBF approach was implemented using radial basis neural network. In general, it consists of two layers: a hidden radial basis layer, and an output linear layer. The transfer function of radial basis neuron is

$$f = \mathrm{e}^{-(||w-p||b)^2} \tag{9}$$

where w is the weight vector of the neuron, p the input vector and b is the bias. Each bias in the first layer is set to 0.8326/spread. This gives radial basis functions that cross 0.5 at weighted inputs of ±spread. It determines the width of an area in the input space to which each neuron responds. Spread should be large enough that neurons respond strongly to overlapping regions of the input space. The second, output layer of linear neurons, calculates the linear combination of outputs from radial basis neurons.

If we present an input vector to a RBF neural network, each neuron in the radial basis layer will output a value according to how close the input vector is to each neuron's weight vector. Radial basis neurons with weight vectors quite different from the input vector p have outputs near zero. These small outputs have only a negligible effect on the linear output neurons. In contrast, a radial basis neuron with a weight vector close to the input vector p produces a value near 1. If a neuron has an output of 1 its output weights in the second layer pass their values to the linear neurons in the second layer. The classification problem is solved successfully if radial neurons are present in the network, which have weights close to data vectors from each class and their linear transformation provides for a response close to the value considered as the identifier of appropriate class [15].

The RBF neural network employed in the study utilized the newrb function implemented in MATLAB. The function newrb iteratively creates a radial basis network by adding

Table 2
Optimum classification performance of RBF neural network when input features were sensors

VOC	Configuration of sensors						
	I	II	III	IV	V	VI	VII
Benzene	14/16	15/16	15/16	16/16	15/16	9/16	15/16
Toluene	15/16	16/16	15/16	16/16	16/16	8/16	16/16
Xylene	16/16	16/16	15/16	15/16	15/16	12/16	16/16
Spread	0.55	1.35	0.12	0.24	0.4	0.24	0.22

one neuron at a time. Neurons are added to the network until the sum-squared error falls beneath an error goal or a maximum number of neurons has been reached [15].

In our network the output layer consisted of one neuron. The network was expected to respond with (−1) for the data vectors associated with benzene, 0 for the data vectors associated with tolune and 1 for xylene.

The RBF was applied to the same data set as did PCA and DFA. Modified leave-one-out cross-validation was used to obtain the training and classification performance for neural network, as implemented before for the DFA validation. In each procedure run, one measurement vector was excluded from 72 element data set and neural network was trained on remaining 71 data points. The excluded data vector was then classified and the accuracy of classification was checked. There were 48 runs of leave-one-out cross-validation procedure as there were 48 non-extreme vectors to be checked if classified properly. Measurements for minimum and maximum concentrations of each compound at four humidity levels (24 in total) were not considered as testing vectors.

It has been reported in literature that once the classifier has been chosen, an optimum classification performance depends on the selection of adequate input features [7]. At first, we investigated the possibility of using sensors responses as input features. We considered following configurations of sensors as inputs to the RBFNN in search for the best classification performance: (I) TGS800, TGS822, TGS824, TGS825, TGS880, and TGS883 (six sensors) (II) TGS800, TGS822, TGS824, TGS825, and TGS880 (five sensors: humidity sensor excluded) (III) TGS822, TGS824, TGS825, and TGS880 (IV) TGS800, TGS824, TGS825, and TGS880 (V) TGS800, TGS822, TGS825, and TGS880 (VI) TGS800, TGS822, TGS824, and TGS880 (VII) TGS800, TGS822, TGS824, and TGS825 (four element combinations out of five non-humidity sensors). A series of simulations were done for every configuration of input vectors searching for the RBF spread, which minimized the number of erroneous classifications. The optimum RBF network performance and associated spread values, when sensors were considered as inputs, are reported in Table 2.

None classification was fully successful when sensor response were used as neural network input features (Table 2). One could observe that humidity sensor impaired classification of benzene, toluene and xylene, as the results improved after removing it from the feature set (I versus others). Therefore, it was not considered further in five and four element sensor configurations. TGS825 appeared to be the key sensor in our classification problem (VI). When removed from feature set, classification performance rapidly broke. Remaining sensor configurations resulted in one or two misclassifications for the whole test data set, which was comparable to DFA performance. Further reduction of input vectors dimensionality, below four features resulted in unacceptable classification errors, so these results were not presented.

Table 3
Optimum classification performance of RBF neural network when inputs were principal components calculated for complete set of sensors (six) or calculated for set of sensors, without humidity sensor (five)

VOC	Configuration of principal components				
	6/6	5/6	4/6	5/5	4/5
Benzene	15/16	16/16	16/16	16/16	16/16
Toluene	16/16	15/16	14/16	16/16	16/16
Xylene	16/16	16/16	16/16	15/16	16/16
Spread	1.50	1.30	1.05	0.95	1.15

The principal components eliminated were those, explaining minimum variance.

As we have not fully succeeded in classification of benzene, toluene and xylene using directly sensor responses for neural network inputs, it was proposed to preprocess the data with feature extraction method and apply the extracted features for neural network inputs. We took advantage of principal components calculated previously and we used them as the RBF network inputs. The search for an adequate set of principal components started from all principal components calculated for a set of sensors including humidity sensor (six components). Then, all principal components calculated for a set of sensors without humidity sensor were used (five components). In the next step, components explaining minimum variance were eliminated from those feature spaces and so on. A series of simulations were done for every configuration of input vectors in search for the RBF spread, which minimized the number of erroneous classifications and results are reported in Table 3.

As listed in Table 3, only one configuration of principal components used as network inputs allowed for 100% success rate in classification of each single compound: benzene, toluene and xylene. It was the case of four principal components obtained from PCA on five gas sensors (4/5). RBF classification performance for the rest of cases shown in Table 3 was comparable to using five or four sensors as neural network input features or applying DFA.

4. Conclusions

The recognition task of three single VOCs: benzene, toluene and xylene in air has been investigated. We have shown that simple measurement system consisting of an

array of commercially available, cheap TGS sensors, coupled with appropriate classifier was capable of solving this classification task. As presented in the article, simple linear discriminator DFA was 100% successful in recognition of benzene, in concentration range 3–73 ppm, whatever the humidity of air. This classifier would be sufficient for the applications targeted at recognition of toxic benzene from less toxic toluene or xylene. In order to accurately recognize each VOC the non linear classifier has been employed. It was shown that RBF neural network faultlessly discriminated each single compound: benzene, toluene and xylene. We showed that in order to reach this classification goal, the defined set of principal components had to be used for a neural network input.

Acknowledgements

Special thanks to the Developers of public domain R statistical software.

References

[1] S.H. Hahn, N. Barsan, U. Weimar, Investigation of CO/CH_4 mixture measured with differently doped SnO_2 sensors, Sens. Actuators B Chem. 78 (2001) 64–68.
[2] M. Schweizer-Berberich, S. Strathmann, U. Weimar, R. Sharma, A. Seube, A. Peyre-Lavigne, W. Gopel, Strategies to avoid VOC cross-sensitivity of SnO_2 based CO sensors, Sens. Actuators B Chem. 58 (1999) 318–324.
[3] M. Ivanovskaya, P. Bogdanov, G. Faglia, P. Nelli, G. Sbervelieri, A. Taroni, On the role of catalytic additives in gas-sensitivity of SnO_2-Mo based thin film sensors, Sens. Actuators B Chem. 77 (2001) 268–274.
[4] R.S. Niranjan, V.A. Chaudhary, S.R. Sinkar, K.R. Patil, K. Vijayamohanan, I.S. Mulla, Surface ruthenated tin oxide thin-film as a hydrocarbon sensor, Sens. Actuators B Chem. 79 (2001) 132–136.
[5] K. Kato, Y. Kato, K. Takamatsu, T. Udaka, T. Nakahara, Y. Matsuura, K. Yoshikawa, Toward the realisation of an intelligent gas sensing system utilizing a non-linear dynamic response, Sens. Actuators B Chem. 71 (2000) 192–196.
[6] R. Menzel, J. Goshnick, Gradient gas sensor microarray for on-line process control, a new dynamic classification model for fast and reliable air quality assessment, Sens. Actuators B Chem. 68 (2000) 115–122.
[7] R. Polikar, R. Shinar, L. Udpa, M.D. Porter, Artificial intelligence method for selection of an optimized sensor array for identification of volatile organic compounds, Sens. Actuators B Chem. 80 (2001) 243–254.
[8] R.E. Shaffer, S.L. Rose-Pehrsson, R.A. McGill, A comparison study of chemical sensor array pattern recognition algorithms, Anal. Chim. Acta 384 (1999) 305–317.
[9] K.L. Goodner, J.G. Dreher, R.L. Rouseff, The dangers of creating false classification due to noise in electronic nose and similar multivariate analyses, Sens. Actuators B Chem. 80 (2001) pp.261–266.
[10] C. Delpha, M. Siadat, M. Lumbreras, Discrimination of refrigerant gas in a humidity controlled atmosphere by using modeling parameters, Sens. Actuators B 62 (2000) 226–232.
[11] A.M. Taurino, C. Distante, P. Sicilano, L. Vasanelli, Quantitative and qualitative analysis of VOCs mixtures by means of microsensors array and different evaluation methods, Sens. Actuators B Chem. 93 (2003) 117–125.
[12] A. Szczurek, P. Malinowski, M. Maciejewska, Application of fuzzy model for identification of selected mixtures of organic compounds, Chemia i Inżynieria Ekologiczna 10 (1999) 1055–1068.
[13] http://doppler.unl.edu/∼bcorner/pca.html ;
[14] A. Webb, Statistical Pattern Recognition, Arnold, New York, 1999, pp. 227–237, 100–132, 140–155.
[15] Matlab, The language of technical computing, Using Matlab5, Release 11, The MathWorks Inc., 1999.
[16] A. Szczurek, M. Maciejewska, Humidity as a discriminative factor in alcohols recognition, Environ. Protect. Eng. 2 (2003) 125–139.
[17] R Development Core Team, R version 1.8.1, Manual, http://www.r-project.org, 2003.
[18] Permissible levels and alarm levels of some substances in ambient air and toleration margins for permissible levels of some substances, Dz. U. 2003.1.12.
[19] B. Walczak, D.L. Massart, Application of radial basis function—partial least squares to non-linear pattern recognition problems: diagnosis of process faults, Anal. Chim. Acta 331 (1996) 187–193.