Measurement 54 (2014) 31–39

Contents lists available at ScienceDirect

Measurement

journal homepage: www.elsevier.com/locate/measurement

A novel sensor selection using pattern recognition in electronic nose

Lei Zhang [a,b,*], Fengchun Tian [a], Guangshu Pei [a]

[a] College of Communication Engineering, Chongqing University, 174 ShaZheng street, ShaPingBa District, Chongqing 400044, China
[b] Department of Computing, The Hong Kong Polytechnic University, Hong Kong

ARTICLE INFO

Article history:
Received 3 November 2013
Received in revised form 4 April 2014
Accepted 8 April 2014
Available online 20 April 2014

Keywords:
Electronic nose
Gas sensors
Gas detection
Sensor selection
Pattern recognition

ABSTRACT

A novel sensor selection using pattern recognition technique in electronic nose (E-Nose) is proposed in this paper. This paper studies the portable E-Nose based on metal oxide semiconductor (MOS) gas sensors for detection of multiple kinds of indoor air contaminants. The characteristics of portability, low cost, multiple targets detection and high performance of E-Nose monitor are the main pursuit for home use. Formaldehyde, benzene, toluene, carbon monoxide, and ammonia are the primary targets of the proposed E-Nose which benefits from the characteristics of the broad spectrum, reproducibility, sensitivity and low-cost of MOS gas sensors. Therefore, a potential and full contribution analysis of the small sized sensor array, in detection of indoor air contaminants coupled with a kernel principal component analysis (KPCA) based linear discriminant analysis (LDA) pattern recognition technique, is presented in this paper. Some experimental findings on the roles of sensors in an E-Nose have also been concluded. The recognition results clearly demonstrate the contribution of each sensor to gas detection which helps the sensor selection in E-Nose design.

1. Introduction

In recent years, electronic nose (E-Nose) with metal oxide semiconductor (MOS) gas sensors or biosensors has been widely studied in different research fields such as air contaminants detection [1–5], food analysis [6–9], medical diagnosis [10–12], and explosion detection [13–15]. The characteristics of broad spectrum response and low-cost of MOS sensors promote the fast development of E-Nose. However, some limitations still exist. For example, one E-Nose can only measure a minority of gases, the portability and the cost are not optimal due to the larger sized (12, 16, 32, etc.) sensor array used in E-Nose. It still doubts that what sized gas sensor array is enough in general use of E-Nose? To make it clearer, the portability and low-cost of E-Nose for home use have to be further studied in real-time monitoring of indoor air quality.

A number of studies and survey demonstrate that formaldehyde (CH_2O), benzene (C_6H_6), toluene (C_7H_8), carbon monoxide (CO), ammonia (NH_3), stimulus (perfume, toiletwater, fruit smell, ethanol), nitrogen dioxide (NO_2), and other VOCs (volatile organic compounds) [16–18] are the principal chemical components that influence the air quality. This paper will focus on the first 11 kinds of chemicals and discriminate them through our designed portable and low-cost E-Nose instrument with a small sized array of MOS gas sensors.

Consider the demands of multi-target contaminants detection and the low-cost of E-Nose, we select four MOS gas sensors from Figaro Inc. in Japan, they are TGS2620, TGS2602, and TGS2201 with two channels (TGS2201A

* Corresponding author at: College of Communication Engineering, Chongqing University, 174 ShaZheng street, ShaPingBa District, Chongqing 400044, China. Tel.: +86 13629788369; fax: +86 23 65103544.
E-mail address: leizhang@cqu.edu.cn (L. Zhang).

http://dx.doi.org/10.1016/j.measurement.2014.04.005
0263-2241/

and TGS2201B). It is known that the appropriate selection of sensor type is very important in electronic nose for various chemicals detection. The criterion for the choice of the four TGS sensors mainly lies in two aspects: first, TGS sensors have characteristics of cost-effective, cross sensitive and spectral response; second, the four sensors selected can detect a wide species such as carbon monoxide, nitric oxide, nitrogen dioxide, ammonia, toluene, ethanol, hydrogen, methane, hydride and VOCs which can basically cover the target contaminants with cross-sensitivity among sensors. Readers can refer to the sensors' datasheets in [19] for details.

With the development of E-Nose study in indoor contaminants measurement, a smaller sized sensor array for detection of more chemicals by an E-Nose with lower cost, higher predictive accuracy and portability becomes the desired objective and main pursuit in instrumental analysis. Therefore, finding of a small sized sensor array for multi-target detection with an E-Nose is meaningful. This paper systematically conducts the analysis of their contribution to E-Nose, proposes a novel sensor selection framework coupled with a KPCA based LDA pattern recognition technique, and proves that the selected small sized sensor array based E-Nose is competent for multiple gases detection.

2. Electronic nose and experiments

2.1. Electronic nose system with embedded MOS gas sensor array

The sensor array in E-Nose system consists of four metal oxide semiconductor gas sensors with TGS series in Figaro Inc. including TGS2602, TGS2620 and TGS2201 with two channels TGS2201A and TGS2201B, which has been introduced in the E-Nose system in previous study [20].

In addition, considering that MOS gas sensors are sensitivity to ambient temperature and humidity, a module with two auxiliary sensors for the temperature (T) and humidity (H) measurement is also used. A 12-bit analog–digital converter is used as the interface between the sensors and the Field Programmable Gate Array (FPGA) processor which can be used for data collection, storage and processing. The E-Nose system is then connected to a personal computer (PC) via a Joint Test Action Group (JTAG) port used for transferring data and debugging programs. The E-Nose system and the experimental platform developed in our laboratory are illustrated in Fig. 1 provided in [21]. The typical response curves of an array in the sampling process with four phases (1. Baseline, 2. Transient response, 3. Steady state response, 4. Recover process) are also plotted in Fig. 1.

2.2. Experiments

From Fig. 1, the gaseous experiments of electronic nose in this paper were employed in the constant temperature and humidity chamber, which can automatically adjust the temperature and humidity by manually setting up. The target gas that was injected to the chamber through a flow meter was collected in a gas bag. A fan is fixed in the chamber for purging the gas and diffusing evenly. Totally, 10 min would be consumed in each experiment and one sample is obtained. The specific experimental procedures including four stages can be totally illustrated as follows

Stage 1: Gas preparing and collection.

Collect each target gas in a bag, and dilute each target gas using pure nitrogen (N_2).

Stage 2: Data collection (major part).

In this stage, there are several steps shown as follows

Step 1: Set the initial temperature and humidity of the chamber. For simulation of the indoor environment, 15, 20, 25, 30 and 35 °C are considered for temperature (T), and 40%, 60% and 80% are considered for relative humidity (RH). Therefore, there are 15 combinations of temperature and humidity considered.

Step 2: Turn on the E-Nose system until the temperature and humidity in the chamber reach the initial setting, and then collect sensor array's responses of baseline for 2 min.

Step 3: Inject target gas by using a flow-meter with time controlled. Then, the sensors will have quick response to target gas and the sensors would reach steady state response after about 8 min. Therefore, one experiment of sample collection would sustain 10 min. The steady state response is extracted as feature to represent the gas texture for pattern analysis.

Stage 3: Air exhaust and chamber cleaning.

After one experiment of sample collection, air exhaust by a pump is necessary for chamber cleaning to recover the sensor response as soon as possible.

Stage 4: Data transferring to PC.

Sensor response data in one experiment is obtained at this time through a JTAG which is connected between the electronic nose and PC. Then the collected data can be transferred to the PC conveniently for data analysis in computer.

Reference concentrations of the first six kinds of gases for all samples can be obtained by specific analysis and instrumental methods. Specifically, the concentration of CH_2O is analyzed using 721G type of spectrophotometer, GC analysis is for C_6H_6, C_7H_8 and NH_3 concentration analysis, and the concentrations of CO and NO_2 are obtained from reference instrument with type of KP826. Consider the needs of indoor contaminates detection, the concentration ranges (the lowest and the highest concentration) in experiments are 0.04–5.32 ppm, 0.17–0.91 ppm, 0.05–0.14 ppm, 5.00–49.0 ppm, 0.09–2.15 ppm and 0.03–1.62 ppm for CH_2O, C_6H_6, C_7H_8, CO, NH_3 and NO_2, respectively.

2.3. Experimental data

In this paper, 11 kinds of chemicals including CH_2O, C_6H_6, C_7H_8, CO, NH_3, NO_2 and stimulus (perfume, toilet-water, fruit smell and ethanol) are studied. All the experiments were employed with respect to the described experimental method. Totally, 721 samples were obtained

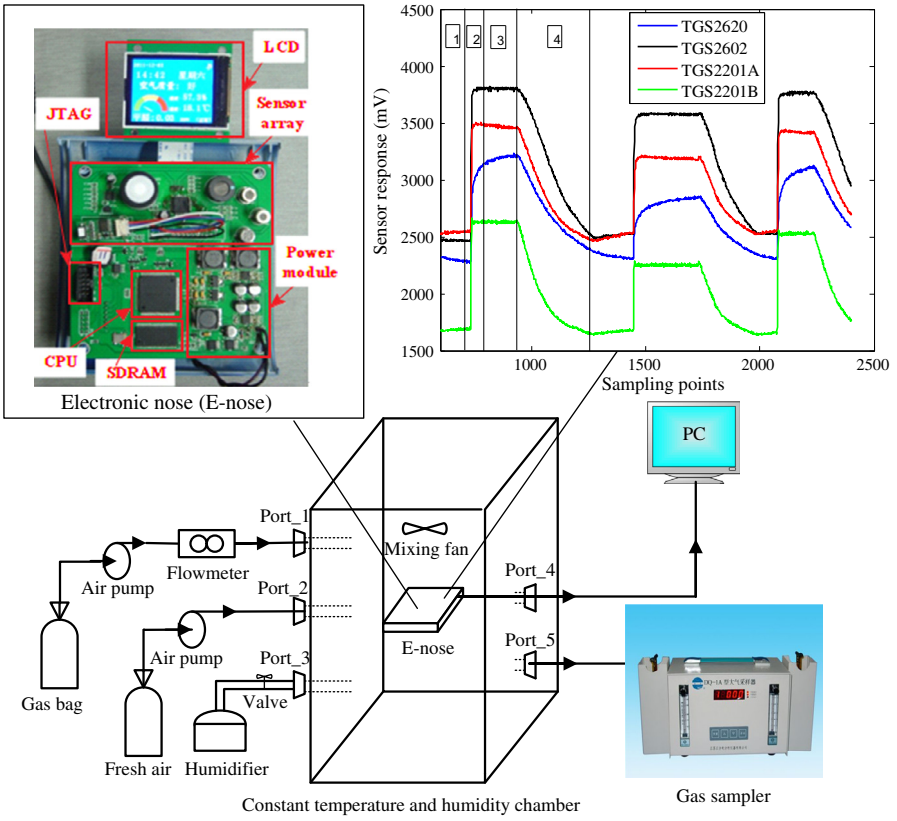

Fig. 1. Portable electronic nose, the experiments and the typical sensor response in this paper.

for the 11 kinds of gases. In pattern analysis, 455 samples are used for learning, and the remaining 266 samples are used for validation. For detail, the concentrations, temperature and relative humidity of the experimental samples have been briefly referred in recent paper [21]. Note that the concentrations of stimulus (perfume, toilet-water, fruit smell and ethanol) are not employed in our experiments.

3. Classification technique

3.1. Kernel based PCA (KPCA)

PCA [22] is an unsupervised method in dimension reduction by projecting correlated variables into another orthogonal feature space and thus a group of new variables with the largest variance (global variance maximization) were obtained. PCA is an effective dimension reduction technique which can preserve the most information of original data. KPCA is a hybrid feature extraction method which consists of kernel transformation and PCA. The original input vectors are mapped to a high dimensional feature space F by using a kernel function which can make the original data linearly separable in the high dimensional space F. The mapping from the original data space to high dimensional feature space can be represented by calculating the symmetrical kernel matrix of input training pattern vectors using a Gaussian kernel function shown as

$$K(\mathbf{x}, \mathbf{x}_i) = \exp\left(\frac{-\|\mathbf{x} - \mathbf{x}_i\|^2}{\sigma^2}\right) \quad (1)$$

where $\mathbf{x}$ and $\mathbf{x}_i$ denotes the observation vectors, σ^2 denotes the kernel parameter.

In general, KPCA is to perform PCA algorithm in the high dimensional feature space and extract nonlinear feature. The dimension size depends on the number of training vectors. In this paper, the kernel principal components of original data are used for learning and recognition in classification tasks with an E-Nose. For easy understanding of KPCA, the specific procedure for calculating the kernel principal components (scores) of training and testing vectors can be described as follows:

Step 1: Computation of the symmetrical Kernel matrix $K_{m\times m}$ of the training vectors, where m denotes the number of training vectors. The element K(i,j) is calculated between vector x_i and x_j (i = 1, ..., m; j = 1, ..., m) using Eq. (1).

Step 2: Regularization of kernel matrix K. Regularization includes two parts: centralization and normalization. The centralization in this paper can be shown by:

$$\mathbf{K} = \mathbf{K} - \frac{1}{m}\mathbf{I}\cdot\mathbf{K} - \frac{1}{m}\mathbf{K}\cdot\mathbf{I} + \frac{1}{m}\mathbf{I}\cdot\mathbf{K}\cdot\mathbf{I} \quad (2)$$

where $\mathbf{I}$ is unit matrix of m by m.

Then, the normalization is realized by

$$\mathbf{K} = \frac{\mathbf{K}}{m} \tag{3}$$

Step 3: Decomposition of K to obtain eigenvalues and eigenvectors. The decomposition is to calculate the equation shown as

$$\mathbf{K} \cdot \mathbf{V} = \lambda \cdot \mathbf{V} \tag{4}$$

where λ and $\mathbf{V}$ denote the eigenvalues and eigenvectors.

Step 4: Determine the final eigenvectors projection_vectors for kernel principal components in classification. Sort the eigenvalues in descending order, and get the sorted new_eigenvalues and the corresponding index used to determine the new_eigenvectors. Then, calculate the accumulative contribution rate (ACR) which is shown by

$$ACR_j = \frac{\sum_{k=1}^{j} new\_eigen\,values_k}{\sum_{i=1}^{m} new\_eigen\,values_i} \times 100, \quad j = 1, \ldots, m \tag{5}$$

where m denotes the number of training vectors and also the number of eigenvalues, j should be calculated to determine the number of kernel principal components for classification, according to the following equation.

$$j = \arg\min\{ACR_j \geqslant CR\} \tag{6}$$

where CR is the given threshold in algorithm, and 95%, 96%, 97%, 98% and 99% are studied in this paper. Then, the **projection_vectors** are determined as the first j column vectors of **new_eigenvectors**.

Step 5: Calculate the kernel principal components KernelPC_training of training vectors as follows:

$$\mathbf{KernelPC\_training} = \mathbf{K} \times \mathbf{projection\_vectors} \tag{7}$$

Step 6: Calculate the kernel matrix Kt of testing vectors. Similar to Step 1, the element Kt(i,j) is calculated between vector xi and xj (i = 1, …, n; j = 1, …, n) using Eq. (1), where n denotes the number of testing vectors.

Step 7: Regularization of kernel matrix Kt which is similar to Step 2, and can be realized using Eqs. (2) and (3).

Step 8: Calculate the kernel principal components KernelPC_testing of testing vectors using the projection_vectors obtained from training vectors in Step 4 shown by:

$$\mathbf{KernelPC\_testing} = \mathbf{Kt} \times \mathbf{projection\_vectors} \tag{8}$$

Step 1–Step 8 illustrate the specific feature extraction method which contains kernel mapping and PCA dimension reduction. In the following, the basic principle of FLDA classifier will be presented to realize the classification.

3.2. FLDA based classifier

FLDA aims to maximize the ratio of the between-class variance and the within-class variance using fisher criterion in any particular data set through a transformation vector $\mathbf{w}$, and therefore promise the maximum separability with a linear decision boundary between two classes.

To a binary classification (two classes), assume the dataset for the two classes to be $\mathbf{X}_1 = \{\mathbf{x}_1^1, \mathbf{x}_1^2, \ldots, \mathbf{x}_1^{N_1}\}$ and $\mathbf{X}_2 = \{\mathbf{x}_2^1, \mathbf{x}_2^2, \ldots, \mathbf{x}_2^{N_2}\}$, respectively, N_1 and N_2 denote the numbers of column vectors for $\mathbf{X}_1$ and $\mathbf{X}_2$, $\{\mathbf{x}_i^j,\ i = 1, 2;\ j = 1, \ldots, N_i\}$ denotes the column vector (observation sample). Then, we set the total dataset $\mathbf{Z}$ in $\mathbf{R}^d$ as $\mathbf{Z} = \{\mathbf{X}_1, \mathbf{X}_2\}$. The description of FLDA is as follows.

The centroid of each class can be calculated by:

$$\boldsymbol{\mu}_i = \frac{1}{N_i} \cdot \sum_{j=1}^{N_i} \mathbf{x}_i^j, \quad i = 1, 2 \tag{9}$$

The within-scatter matrix $\mathbf{S}_i$ of class i is

$$\mathbf{S}_i = \sum_{j=1}^{N_i} (\mathbf{x}_i^j - \boldsymbol{\mu}_i)(\mathbf{x}_i^j - \boldsymbol{\mu}_i)^{\mathrm{T}}, \quad i = 1, 2 \tag{10}$$

where symbol T denotes transpose.

The within-class scatter matrix $\mathbf{S}_w$ and the between-class scatter matrix $\mathbf{S}_b$ can be described as

$$\mathbf{S}_w = \mathbf{S}_1 + \mathbf{S}_2 \tag{11}$$

$$\mathbf{S}_b = \sum_{i=1}^{2} N_i \cdot (\boldsymbol{\mu}_i - \bar{\mathbf{Z}}) \cdot (\boldsymbol{\mu}_i - \bar{\mathbf{Z}})^{\mathrm{T}} \tag{12}$$

where $\bar{\mathbf{Z}}$ denotes the centroid of the total dataset $\mathbf{Z}$.

To obtain the optimal transformation vector $\mathbf{w}^*$, the fisher criterion in terms of $\mathbf{S}_w$ and $\mathbf{S}_b$ is expressed as

$$J(\mathbf{w}) = \mathbf{w}^{\mathrm{T}}\mathbf{S}_b\mathbf{w} / \mathbf{w}^{\mathrm{T}}\mathbf{S}_w\mathbf{w} \tag{13}$$

where $\mathbf{w}$ is the transformation vector in a binary classification and the optimal $\mathbf{w}^*$ is represented as

$$\mathbf{w}^* = \arg\max\{J(\mathbf{w})\} \tag{14}$$

The $\mathbf{w}^*$ can be calculated by solving the eigenvalue problem. In this paper, $\mathbf{w}^*$ is the eigenvector which corresponds to the largest eigenvalue of matrix $\mathbf{S}_w^{-1}\mathbf{S}_b$. Note that, to a multi-class problem, $\mathbf{w}^*$ should be a matrix.

Then, the unknown observation vector $\mathbf{x}$ can be transformed through $\mathbf{w}^*$ according to

$$\mathbf{x}' = \mathbf{x} \cdot \mathbf{w}^* \tag{15}$$

Finally, the Euclidean distance is used in the decision of $\mathbf{x}$

$$\text{if } \|\mathbf{x}' - \boldsymbol{\mu}_1\|_2 < \|\mathbf{x}' - \boldsymbol{\mu}_2\|_2, \quad \mathbf{x} \in \{\text{class 1}\}; \quad \text{otherwise}, \mathbf{x} \in \{\text{class 2}\} \tag{16}$$

where symbol $\|\cdot\|_2$ denotes the l_2-norm, $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$ denote the center vectors of class 1 and class 2, respectively.

3.3. Multi-class classification with an E-Nose

This paper studies the discrimination of $k = 6$ kinds of pollutant gases based on one-against-one (OAO) strategy [23]. Therefore, this paper will build one sub-classifier between two arbitrary classes, and totally $k \cdot (k-1)/2 = 15$

binary classifiers will be produced using OAO strategy. Furthermore, a majority voting mechanism, in which the class with the most votes would be the discriminated class of an unknown sample vector in decision level, is used. The classification method of FLDA with two classes has been presented in Section 3.2. The multi-class classification of FLDA is to perform the same procedure between two arbitrary classes based on OAO strategy first, and then vote on the results of each sub-classifier.The diagram of the proposed classification method with an E-Nose has been illustrated in Fig. 2, wherein PART I (the top part) represents the integral flowchart of the classifications. The region with red and dashed line is the learning process of KPCA, the region with blue and dashed line is the FLDA training module and the region with green and dashed line is the FLDA testing module in classification; PART II (the middle part) presents the internal structure with OAO strategy of FLDA training module for obtaining the transformation matrix $\mathbf{W}^*$; PART III (the bottom part) presents the internal structure of FLDA testing module for decision (classification) of unknown sample. All the algorithms in this paper are implemented in the platform of Matlab software.

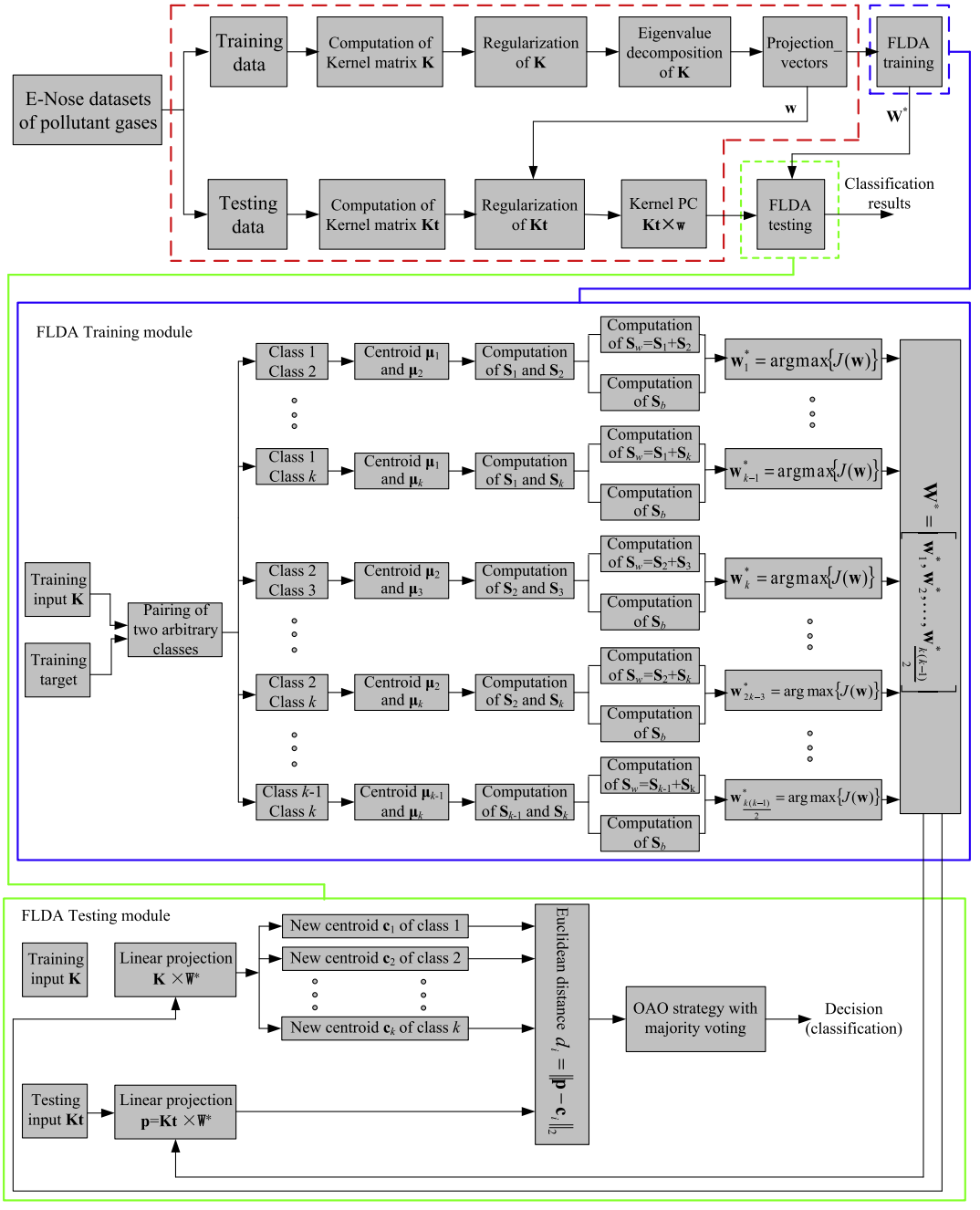

Fig. 2. Diagram of the KPCA-FLDA classification technique.

4. Results and discussion

This paper further presents the contribution analysis of selective MOS gas sensors to low-cost electronic nose in recognition of indoor air contaminates. Therefore, with respect to different combinations of TGS2620, TGS2602, TGS2201A and TGS2201B, 15 cases are studied totally and the contributions of sensors have also been analyzed.

Table 1 presents the classification accuracies of each gas with different sensors considered. "√"denotes that the sensor is considered; "×"denotes that the sensor is not considered. From Table 1, we can qualitatively analyze the contribution of each sensor to the gas classification by an E-Nose. From case 1 to case 4, there is only one sensor used in pattern recognition; from case 5 to case 10, there are two sensors used; from case 11 to case 14, 3 sensors are used in detection; in case 15, totally 4 sensors are used in E-Nose. It is obvious that the classification accuracies are improved with the increasing number of sensors. From Table 1, we can find that TGS2620 contributes little to formaldehyde (see case 11), while TGS2602 and TGS2201A play a more important role to formaldehyde detection (see case 8). For benzene, the four sensors have no obvious roles in detection. TGS2620 and TGS2201A

Table 1
Classification accuracy of multiple kinds of indoor contaminants in 15 cases with environmental factors considered.

Case	Classification accuracy (%)							MOS gas sensors considered			
	CH_2O	C_6H_6	C_7H_8	CO	NH_3	NO_2	Stimulus	TGS2620	TGS2602	TGS2201A	TGS2201B
1	57.1	33.3	22.7	90.0	45.0	76.9	47.2	×	×	×	√
2	80.9	33.3	22.7	60.0	40.0	30.8	38.7	×	×	√	×
3	52.4	33.3	18.2	85.0	60.0	46.2	60.4	×	√	×	×
4	50.8	54.1	31.8	90.0	35.0	7.69	54.7	√	×	×	×
5	90.5	45.8	81.8	90.0	55.0	53.8	65.1	×	×	√	√
6	84.1	50.0	40.9	85.0	65.0	69.2	68.9	×	√	×	√
7	95.2	58.3	27.3	75.0	85.0	53.9	77.4	×	√	√	×
8	52.4	33.3	63.6	80.0	65.0	0.00	82.1	√	×	×	√
9	76.2	62.5	68.2	90.0	40.0	38.5	54.7	√	×	√	×
10	73.0	45.8	36.4	90.0	80.0	46.2	83.0	√	√	×	×
11	98.4	50.0	77.3	90.0	75.0	84.6	85.9	×	√	√	√
12	90.5	66.7	81.8	80.0	95.0	53.9	83.0	√	×	√	√
13	84.1	66.7	63.6	90.0	85.0	69.2	89.6	√	√	×	√
14	93.7	70.8	90.9	90.0	80.0	76.9	87.7	√	√	√	×
15	98.4	75.0	86.4	95.0	90.0	69.2	95.3	√	√	√	√

Note: "√"denotes that the sensor is considered; "×"denotes that the sensor is not considered.

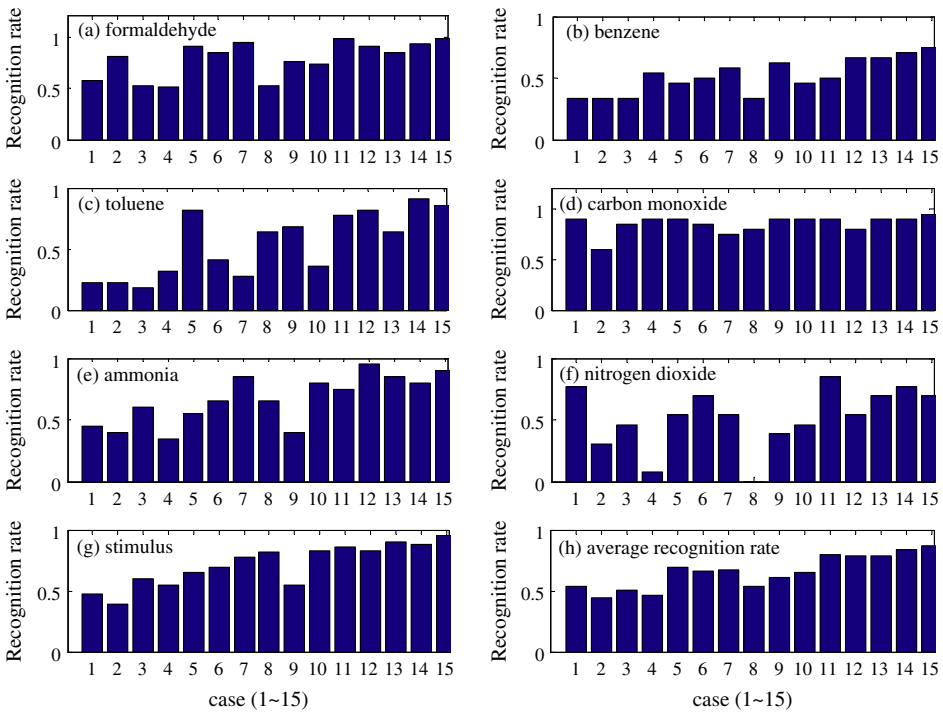

Fig. 3. Variations of recognition rate for each analyte in 15 cases.

contribute very well to toluene detection (see cases 3, 7 and 10). The recognition result of CO seems to be more optimistic for each sensor except the case 2. From the result of ammonia, we can see that TGS2602 has little contribution (see case 12). For NO_2, the recognition results are not good for every case. The reason may result from the small number of experimental samples of NO_2, which is less than number of other gases samples. From the results of stimulus, we can see that the four sensors contribute equally in detection.

For visually observe the variations of classification accuracy of each gas with sensor reduction, Fig. 3 with bar plots of gas recognition rate in the 15 cases is presented, wherein Fig. 3(h) illustrates the average recognition rate of gases in each case. From Fig. 3 we can clearly find the obvious fluctuations of recognition rates of C_6H_6, C_7H_8, CO, NO_2, NH_3 and stimulus in 15 cases. The contribution of each sensor to each chemical can be judged positively or negatively in such qualitative way.

It is worth noting that the classifications presented in Table 1 are obtained with environmental variables (i.e. temperature and humidity) considered. In order to show the importance of environmental factors, we presented the classification results without considering temperature and humidity sensor variables in Table 2. Also, the same 15 cases are studied. By comparing the 15th case between

Table 2
Classification accuracy of multiple kinds of indoor contaminants without considering environmental factors in 11 cases.

Case	Classification accuracy (%)							MOS gas sensors considered			
	CH_2O	C_6H_6	C_7H_8	CO	NH_3	NO_2	Stimulus	TGS2620	TGS2602	TGS2201A	TGS2201B
1	4.76	37.5	27.2	45.0	25.0	53.9	20.8	×	×	×	√
2	31.8	37.5	27.3	30.0	50.0	23.1	30.2	×	×	√	×
3	17.5	87.5	0.00	30.0	0.00	69.2	45.3	×	√	×	×
4	36.5	54.2	63.6	75.0	10.0	7.69	21.7	√	×	×	×
5	12.7	58.3	59.1	40.0	55.0	53.9	50.0	×	×	√	√
6	26.9	45.8	40.9	50.0	30.0	84.6	72.6	×	√	×	√
7	57.1	75.0	4.55	30.0	75.0	46.2	53.8	×	√	√	×
8	44.4	45.8	77.3	100.0	35.0	23.1	66.9	√	×	×	√
9	74.6	37.5	63.6	60.0	15.0	23.1	35.9	√	×	√	×
10	55.6	4.17	54.6	70.0	60.0	69.2	69.8	√	√	×	×
11	63.5	70.8	68.2	35.0	90.0	69.2	70.8	×	√	√	√
12	57.1	62.5	54.6	90.0	75.0	46.2	59.4	√	×	√	√
13	73.0	45.8	63.6	75.0	20.0	84.6	94.3	√	√	×	√
14	82.5	54.2	95.5	55.0	90.0	53.9	83.9	√	√	√	×
15	92.1	58.3	90.9	60.0	95.0	100.0	89.6	√	√	√	√

Note: "√"denotes that the sensor is considered; "×"denotes that the sensor is not considered.

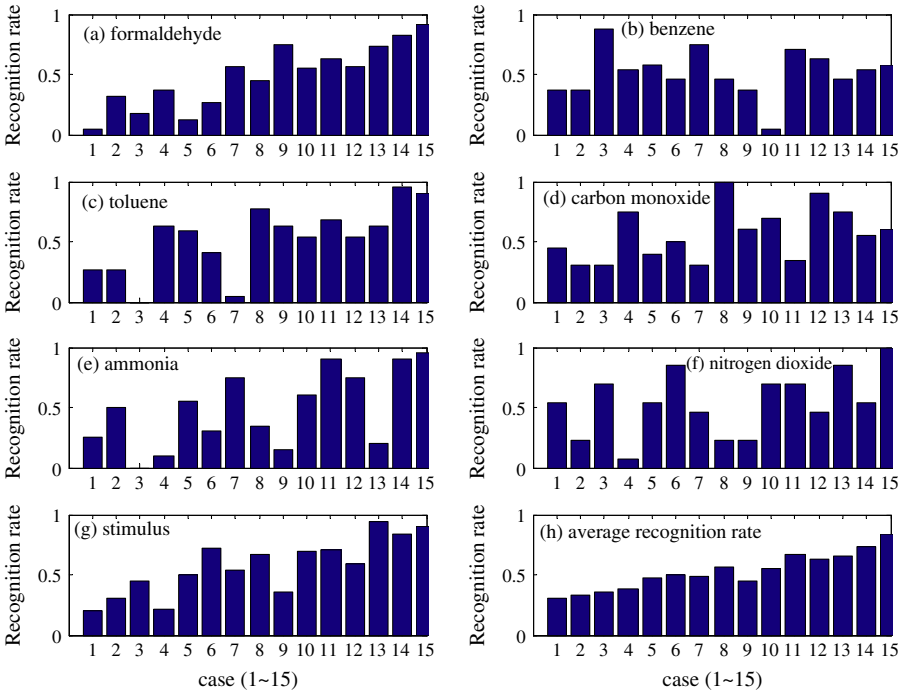

Fig. 4. Variations of recognition rate for each analyte in 15 cases without considering the environmental factors (temperature and humidity).

Tables 1 and 2, we can see that the ambient factors (temperature and humidity) have positive effect in E-Nose detection. Similarly, the bar plots of the recognition rates of each gas in 15 cases have been visualized in Fig. 4 wherein Fig. 4(h) illustrates the average recognition rate of seven gases in each case.

Through the contribution analysis of 4 MOS gas sensors in E-Nose monitor, we can conclude that the selected sensor array of TGS2620, TGS2602, TGS2201A and TGS2201B are feasible for monitoring of the common indoor air contaminants by a low-cost E-Nose instrument. The characteristics of low-cost, portability and high performance (classification accuracy and multiple target detections) with only 3 MOS gas sensors are very persuasive in this paper.

The quantitative estimation of the gases concentration should be the fundamental task of E-Nose in most application. To discuss the possibility and suitability of the E-Nose in quantitative application, we will focus on the algorithmic aspect. For concentration estimation, regression models (e.g. artificial neural network) between the sensor response of an array and the target concentration can be constructed. In our recent work [21], we have presented a further analysis in quantitative application of the E-Nose based on multilayer perceptron neural network and intelligent optimization algorithms for concentration estimation of six contaminants. The estimation results demonstrate that the proposed low-cost E-Nose is feasible in real application. Besides, considering the implicit problems of metal oxide semiconductor based sensors, for example, drift, environmental interferences and weak reproducibility, we have developed effective algorithms to address these issues, such as drift prediction [24], interference elimination [25] and reproducibility enhancement [26] in large scale application, etc. Currently, the E-Nose prototypes have been developed for actual application by different customers.

5. Conclusions

This paper aims to find the potential of E-Nose with a small sized sensor array in various detections. A novel sensor selection method is proposed using KPCA based LDA pattern recognition technique in E-Nose. A potential contribution analysis of TGS2620, TGS2602 and TGS2201 sensors in low-cost electronic nose monitoring for multiple indoor air contaminants, i.e. formaldehyde, benzene, toluene, ammonia, carbon monoxide and stimulus, is first presented. The characteristic of low-cost, portability and high performance is the main pursuit in our E-Nose, and experimental results fully prove the feasibility in usage of only a small sized sensor array of four MOS gas sensors in E-Nose. In addition, we also show that the environmental factors are also important in classification of indoor air contaminants by an E-Nose qualitatively.

Acknowledgements

We would like to express our sincere appreciation to the editor and the anonymous reviewers for their insightful comments. This work was supported by Hong Kong Scholar Program (No. XJ2013044), and also funded by China Postdoctoral Science Foundation (No. 2014M550457).

References

[1] S. De Vito, E. Massera, M. Piga, L. Martinotto, G. Di Francia, On field calibration of an electronic nose for benzene estimation in an urban pollution monitoring scenario, Sens. Actuators B 129 (2008) 750–757.
[2] S.M. Kanan, O.M. El-Kadri, I.A. Abu-Yousef, M.C. Kanan, Semiconducting metal oxide based sensors for selective gas pollutant detection, Sensors 9 (2009) 8158–8196.
[3] K.A. Ngo, P. Lauque, K. Aguir, High performance of a gas identification system using sensor array and temperature modulation, Sens. Actuators B 124 (2007) 209–216.
[4] L. Dang, F. Tian, L. Zhang, C. Kadri, X. Yin, X. Peng, S. Liu, A novel classifier ensemble for recognition of multiple indoor air contaminants by an electronic nose, Sens. Actuators A 207 (2014) 67–74.
[5] L. Zhang, F. Tian, A new kernel discriminant analysis framework for electronic nose recognition, Anal. Chim. Acta 816 (2014) 8–17.
[6] L. Gil-Sánchez, J. Soto, R. Martínez-Máñez, E. Garcia-Breijo, J. Ibáñez, E. Llobet, A novel humid electronic nose combined with an electronic tongue for assessing deterioration of wine, Sens. Actuators A 171 (2) (2011) 152–158.
[7] A. Berna, Metal oxide sensors for electronic noses and their application to food analysis, Sensors 10 (2010) 3882–3910.
[8] N. El Barbri, E. Llobet, N. El Bari, X. Correig, B. Bouchikhi, Electronic nose based on metal oxide semiconductor sensors as an alternative technique for the spoilage classification of red meat, Sensors 8 (2008) 142–156.
[9] S. Labreche, S. Bazzo, S. Cade, E. Chanie, Shelf life determination by electronic nose: application to milk, Sens. Actuators B 106 (2005) 199–206.
[10] A.D. Wilson, M. Baietto, Advances in electronic-nose technologies developed for biomedical applications, Sensors 11 (2011) 1105–1176.
[11] A. D'Amico, G. Pennazza, M. Santonico, E. Martinelli, C. Roscioni, G. Galluccio, R. Paolesse, C. Di Natale, An investigation on electronic nose diagnosis of lung cancer, Lung cancer 68 (2) (2010) 170–176.
[12] V.H. Tran, P.C. Hiang, M. Thurston, P. Jackson, C. Lewis, D. Yates, G. Bell, P.S. Thomas, Breath analysis of lung cancer patients using an electronic nose detection system, IEEE Sens. J. 10 (9) (2010) 1514–2010.
[13] K. Brudzewski, S. Osowski, W. Pawlowski, Metal oxide sensor arrays for detection of explosives at sub-parts-per million concentration levels by the differential electronic nose, Sens. Actuators B 161 (2012) 528–533.
[14] J.W. Gardner, Review of Conventional Electronic Noses and their Possible Application to the Detection of Explosives, Springer, New York, 2004.
[15] J. Yinon, Peer reviewed: detection of explosives by electronic noses, Anal. Chem. 75 (2003) 99–105.
[16] A.P. Jones, Indoor air quality and health, Atmos. Environ. 33 (1999) 4535–4564.
[17] K. Sakai, D. Norbäck, Y. Mi, E. Shibata, M. Kamijima, T. Yamada, Y. Takeuchi, A comparison of indoor air pollutants in Japan and Sweden: formaldehyde, nitrogen dioxide, and chlorinated volatile organic compounds, Environ. Res. 94 (2004) 75–85.
[18] S.C. Lee, M. Chang, Indoor and outdoor air quality investigation at schools in Hong Kong, Chemosphere 41 (2000) 109–113.
[19] http://www.figaro.co.jp/en/product/index.php?mode=search&kbn=1.
[20] L. Zhang, F.C. Tian, H. Nie, L.J. Dang, G. Li, Q. Ye, C. Kadri, Classification of multiple indoor air contaminants by an electronic nose and a hybrid support vector machine, Sens. Actuators B 174 (2012) 114–125.
[21] L. Zhang, F.C. Tian, Performance study of multilayer perceptrons in a low-cost electronic nose, IEEE Trans. Instrum. Meas., 2014, doi.10.1109/TIM.2014.2298691 (in press).
[22] J. Karhunen, Generalization of principal component analysis, optimization problems and neural networks, Neural Network 8 (1995) 549–562.

[23] L.H. Chiang, M.E. Kotanchek, A.K. Kordon, Fault diagnosis based on Fisher discriminant analysis and support vector machines, Comput. Chem. Eng. 28 (2004) 1389–1401.
[24] L. Zhang, F.C. Tian, S. Liu, L. Dang, X. Peng, X. Yin, Chaotic time series prediction of E-Nose sensor drift in embedded phase space, Sens. Actuators B 182 (2013) 71–79.
[25] L. Zhang, F.C. Tian, L. Dang, G. Li, X. Peng, X. Yin, S. Liu, A novel background interferences elimination method in electronic nose using pattern recognition, Sens. Actuators A 201 (2013) 254–263.
[26] L. Zhang, F.C. Tian, X. Peng, X. Yin, A rapid discreteness correction scheme for reproducibility enhancement among a batch of MOS gas sensors, Sens. Actuators A 205 (2014) 170–176.