Sensors and Actuators A 207 (2014) 67–74

Contents lists available at ScienceDirect

Sensors and Actuators A: Physical

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

A novel classifier ensemble for recognition of multiple indoor air contaminants by an electronic nose

Lijun Dang[a], Fengchun Tian[a], Lei Zhang[a,*], Chaibou Kadri[a], Xin Yin[a], Xiongwei Peng[a], Shouqiong Liu[b]

[a] College of Communication Engineering, Chongqing University, 174 ShaZheng Street, ShaPingBa District, Chongqing 400044, China
[b] Academy of Metrology and Quality Inspection, Chongqing 401123, China

ARTICLE INFO

Article history:
Received 5 July 2013
Received in revised form
17 December 2013
Accepted 18 December 2013
Available online 27 December 2013

Keywords:
Classifiers ensemble
Electronic nose
Multi-class recognition
Feature extraction
Support vector machine

ABSTRACT

This paper presents a novel multiple classifiers system called as improved support vector machine ensemble (ISVMEN) which solves a multi-class recognition problem in electronic nose (E-nose) and aims to improve the accuracy and robustness of classification. The contributions of this paper are presented in two aspects: first, in order to improve the accuracy of base classifiers, kernel principal component analysis (KPCA) method is used for nonlinear feature extraction of E-nose data; second, in the process of establishing classifiers ensemble, a new fusion approach which conducts an effective base classifier weighted method is proposed. Experimental results show that the average classification accuracy has been improved from less than 86% to 92.58% compared with that of base classifiers. Besides, the proposed fusion method is also superior to MV fusion method (majority voting) which has 90.1% of classification accuracy. Especially, the proposed ISVMEN can obtain the best discrimination accuracy for C_7H_8, CO and NH_3, almost 100% classification accuracy was obtained using our method. Therefore, it is easy to come to the conclusion that, in average, the proposed method is better significantly than other methods in classification and generalization performance.

1. Introduction

Nowadays, environmental pollution is one of the most critical concerns for governments and individuals. Consequently, there is a resurgence of interest in developing measurement techniques for air quality monitoring. Owing to their portability, real-time operability and ease of use, air quality monitors based on E-nose (that generally combines sensor array with intelligent pattern recognition technology) have attracted consumer's affection [1,2]. However, the performance of these instruments depends extremely on the pattern recognition scheme in use. This paper aims at proposing an effective multi-class recognition model for discrimination of multiple indoor air contaminants. But, in the process of establishing classification model, two main issues need to be considered. That is, classification accuracy and robustness. Specifically, an ideal classifier should not only be able to identify the targets accurately but also tolerate environmental noise (interferences). The general classification methods cannot meet these requirements very well. Fortunately, it has been shown that classifiers ensemble approach can improve both the prediction precision [3] and generalization performance of a recognition system. Moreover, such kinds of ensemble approaches have been widely used in the multi-class recognition problems [4,5]. Ensemble method is a learning approach that many models are combined to solve a given problem. It has been proved in improving the generalization performance of individual models (or base models). The provided base models should be accurate enough and error-independent (diverse) in their predictions.

Although ensemble method is one of the advanced pattern recognition techniques within the machine learning community, only a few studies on their applications in E-nose data processing have been reported in the literatures. Bagging decision trees were used for E-nose applications and their VLSI implementation using 3D chip technology was reported [6]. Shi et al. [7,8] used heterogeneous classifiers including density models, KNN, ANN and SVM for odor discrimination. In [9], the authors showed how ensemble learning methods could be used in an array of chemical sensors (non-selective field transistors) to cope with the interference problem. Gao et al. [10,11] used modular neural networks ensemble to predict simultaneously both the classes and concentrations of several kinds of odors; in the first approach, they used MLPs for base learners, and in their second approach a module or panel comprises various predictors namely MLPs MVLR, QMVLR, and SVM were used. In [12], Hirayama et al. demonstrated that it was possible to detect liquid petrol gas (LPG) calorific power with high recognition rate (up to 99%) using an E-nose and a committee of machines,

[*] Corresponding author. Tel.: +86 13629788369; fax: +86 23 65103544.
E-mail address: leizhang@cqu.edu.cn (L. Zhang).

0924-4247/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.sna.2013.12.029

even with the failure (fault) of one sensor. In [13], Bona et al. used a hybrid algorithm to generate an ensemble of 100 multi-layered perceptrons (MLPs) for the classification of seven categories of coffee. Recently, Vergara et al. [14] proposed an ensemble method that used support vector machines as base classifiers to cope with the problem of drift in chemical gas sensors. Amini et al. [15] used an ensemble of classifiers on data from a single metal oxide gas sensor (SP3-AQ2, FIS Inc., Japan) operated at six different rectangular heating voltage pulses (temperature modulation), to identify three gas analytes including methanol, ethanol and 1-butanol at range of 100–2000 ppm.

Feature extraction is one of the key steps in pattern recognition systems. Principal component analysis (PCA) and independent component analysis (ICA) are among the most widely used feature extraction methods. However, both PCA and ICA are linear feature extraction methods; they may therefore become ineffective in case of nonlinear features. On the other hand, kernel principal component analysis (KPCA) is a nonlinear feature extraction method that integrates kernel trick into standard PCA [16]. It does feature extraction by mapping the original inputs into a high-dimensional feature space, and then the new features are analyzed by PCA in the high-dimensional feature space [17]. Among the above mentioned three methods, KPCA was found to have a better performance in nonlinear feature extraction [18].

In real-time applications, due to varying atmospheric conditions (i.e. temperature, humidity, pressure, etc.) the sensors' responses also vary nonlinearly with gas concentration. The nonlinearity of the sensor array can adversely affect the precision and robustness of the classifier. Therefore, there is a need to come up with an alternative that takes this aspect into consideration. This paper proposes a novel classifiers ensemble which combines KPCA (for feature extraction) and SVM for classification of multiple indoor air pollutants. Firstly, KPCA method was used to extract separable features from the E-nose data; then five SVM base classifiers were trained using different training samples. Finally, the outputs from these base classifiers were combined using an effective weight fusion method.

2. Experiment

The datasets used in this paper were obtained by our E-nose system. Detailed description of the E-nose system can be found in our previous publications [19,20]. However, to make the paper self-contained, we reproduce the system structure and describe briefly the experimental setup.

2.1. Electronic nose system

The sensor array in our E-nose system comprises 7 sensors: four metal oxide semi-conductor gas sensors (TGS2620, TGS2602, TGS2201A, and TGS2201B), humidity, temperature and oxygen. A 12-bit analog-digital converter (A/D) is used as interface between the sensor array and a Field Programmable Gate Array (FPGA) processor. The A/D converts analog signals from sensor array into digital signals which are used by the FPGA for further processing. The FPGA also serves as a control unit. Data collected from the sensor array can be saved as ASCII text files on a PC through JTAG (Joint Test Action Group) port and related software.

2.2. Experimental setup

The experimental platform mainly consists of E-nose system, PC, temperature–humidity controlled chamber, humidifier, air sampler, flow meter, air pump, standard instrument and so on. The specific experimental setup [19] is shown in Fig. 1. The experimental setup has also been mentioned in [21,22]. All experiments were carried out inside the chamber. The experimental procedure in this paper can be summarized as follows: firstly, set the required temperature and humidity control after placing the E-nose system in temperature–humidity chamber; then inject the target gas into the chamber using a pump; finally, the E-nose system will be exposed to the target gas and begin to collect data, furthermore, an air sampler will be used for sampling gases. It is worth noting that a single experiment consists of three stages: exposure to clean air for 2 min for baseline; exposure to gas analyte for 8 min for response; exposure to clean air for 2 min again to allow the sensors recover.

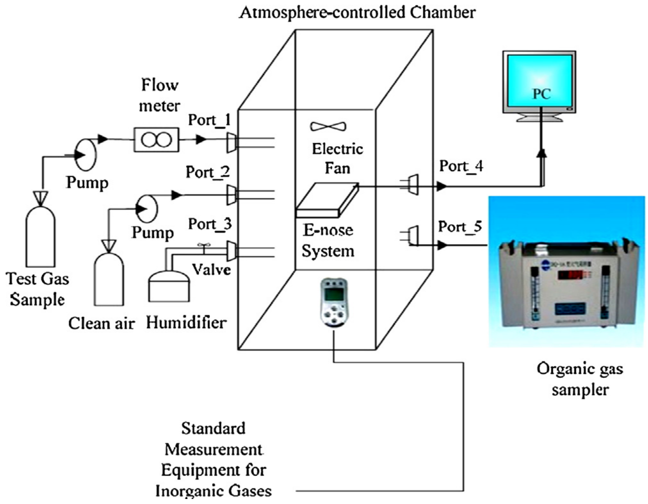

Fig. 1. Experimental platform of E-nose.

The sensor array in this work is composed of four metal oxide semi-conductor gas sensors (i.e. TGS2602, TGS2201A, TGS2201B, and TGS2620) due to their high sensitivity and quick response to detectable gases. The sensing element of the sensors is comprised of a metal oxide semiconductor layer formed on an alumina substrate of a sensing chip together with an integrated heater. In the presence of a detectable gas, the sensor's conductivity increases depending on the gas concentration in the air. Fig. 2 illustrates the sensor response process when exposed to target gas in an experiment. In our experiment, the sampling frequency is set to 1/6 Hz and thus 20 observations would be obtained per minute. Then it's clear to see that the experiment involved 8 min of the exposure of the sensors to gas, and 2 min for the sensors to recover.

2.3. E-nose data

This paper contributes to the classification of six indoor air contaminants including formaldehyde (HCHO), benzene (C_6H_6),

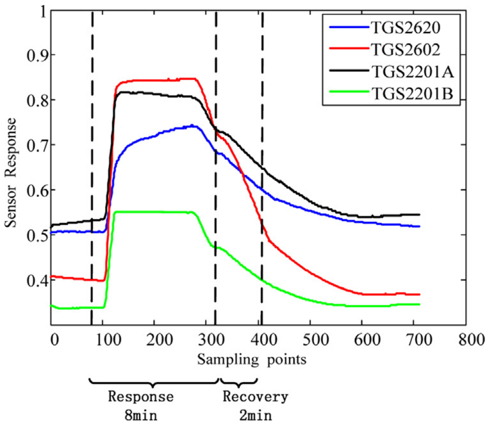

Fig. 2. Response of sensors array to formaldehyde.

Table 1
Distribution of training, test and validation sets.

Data sets	Number of samples in the subset					
	HCHO	C_6H_6	C_7H_8	CO	NH_3	NO_2
dataset_tr	156	99	40	35	29	18
dataset_te	52	32	13	11	9	6
dataset_va	52	33	13	12	10	6
Total	260	164	66	58	48	30

toluene (C_7H_8), carbon monoxide (CO), ammonia (NH_3) and nitrogen dioxide (NO_2). To simulate the indoor environments (e.g. temperature and humidity), the samples for each contaminant were obtained under certain temperature–humidity conditions: 15, 25, 30, and 35 °C for temperature, 40%, 60%, and 80% for relative humidity. Actually, it is nearly impossible to go through all of the temperature and humidity. How to ensure the precision of the product at a different temperature/humidity seems to be especially tricky. In other words, the generalization performance of the proposed classifier is of great importance. Ensemble method has been used to solve this problem in the current work. As to the pressure, all of the experiments were implemented at standard atmospheric pressure for that our product is used in indoor environment. In addition, with the limits of the experimental facilities, we do not change the air pressure conditions as the temperature/humidity does.

The numbers of samples for HCHO, C_6H_6, C_7H_8, CO, NH_3, and NO_2 are 260, 164, 66, 58, 48 and 30 respectively. Detailed information on these samples is available in our previous publication [19].

During the process of establishing ISVMEN model, the whole dataset was divided into three parts: training set, testing set, and validation set. Firstly, 20% of the samples were labeled as **dataset_va** which were obtained by using Kennard–Stone sequential (KSS) algorithm used in [23] and the 80% remaining samples were labeled as **dataset_learning**. Then we randomly select 75% samples from **dataset_learning** to form the training set labeled as **dataset_tr**, while the remaining 25% are left for test labeled as **dataset_te**. We should implement repeated random samplings for L times if there are L base classifiers in the ensemble model. Finally, **dataset_tr** and **dataset_te** were used to generate base classifiers and the corresponding weight of each classifier; **dataset_va** was used to validate the performance of the ensemble. The distribution of training set (**dataset_tr**), test set (**dataset_te**), and validation set (**dataset_va**) for each class is shown in Table 1.

Although the experimental setup does not change too much compared with the previous papers [19], the classification model for indoor air pollutants proposed in this work is quite different. In [19], the proposed classification model HSVM is based on single classifier and linear feature extraction method. However, in this work, some innovations have been made to improve the prediction precision and generalization performance of the classifier. The contributions of this paper are described as follows. First, in order to improve the accuracy of base classifier, KPCA method is used to extract nonlinear feature of six different indoor air pollutants; second, in the process of establishing classifiers ensemble, a new fusion approach that employs an effective weighted method for classifier ensemble is proposed.

3. Methodology

3.1. KPCA algorithm for feature extraction

PCA and ICA have been successfully used for feature extraction in E-nose systems. In this section, a nonlinear formulation of PCA is described. For in-depth description of PCA and ICA we refer the reader to [24,25].

KPCA is one approach of generalizing linear PCA into nonlinear case using the kernel method. The key idea of KPCA is through the nonlinear transform $\Phi(\cdot)$ which maps the sample data from the input space to high-dimensional space F, and performs PCA in the new feature space [26–29].

Suppose that there are M observation samples $\mathbf{X}_i$ with $\mathbf{x}_i \in R^N (i = 1, 2, \ldots, M)$, $\Phi(\mathbf{x}_i)$ represents a high-dimensional feature space and N denotes the length of the original $\mathbf{x}_i$, then if these vectors $\Phi(\mathbf{x}_i)$ meet the zero mean condition, the covariance matrix in the new feature space can be expressed as [30]:

$$\mathbf{C} = \frac{1}{M}\sum_{i=1}^{M}\Phi(\mathbf{x}_i)\Phi(\mathbf{x}_i)^{\mathrm{T}} \tag{1}$$

The eigenvalue decomposition of covariance matrix $\mathbf{C}$ is given by

$$\lambda\boldsymbol{\upsilon} = \mathbf{C}\boldsymbol{\upsilon} \tag{2}$$

where λ and $\boldsymbol{\upsilon}$ denote the eigenvalue and eigenvector of $\mathbf{C}$, respectively.

The eigenvector $\boldsymbol{\upsilon}$ can be expressed in terms of linear combination of the samples

$$\boldsymbol{\upsilon} = \sum_{i=1}^{M}\beta_i\Phi(\mathbf{x}_i) \tag{3}$$

Taking into consideration our mapping $\Phi(\mathbf{x}_k)$ and the result in Eq. (2), we can write

$$\lambda(\Phi(\mathbf{x}_k)\cdot\boldsymbol{\upsilon}) = \Phi(\mathbf{x}_k)\cdot\mathbf{C}\boldsymbol{\upsilon} \quad (k = 1, 2, \ldots, M) \tag{4}$$

Let's define a $M \times M$ matrix $\mathbf{K}$ as

$$\mathbf{K}_{ij} = \mathrm{K}(\mathbf{x}_i, \mathbf{x}_j) = \langle\Phi(\mathbf{x}_i), \Phi(\mathbf{x}_j)\rangle \quad (i, j = 1, 2, \ldots, M) \tag{5}$$

Combining Eqs. (3)–(5), we can obtain the following equation

$$M\lambda\boldsymbol{\alpha} = \mathbf{K}\boldsymbol{\alpha} \tag{6}$$

where $M\lambda$ is the eigenvalue of $\mathbf{K}$, coefficient vector $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \ldots, \alpha_M)^{\mathrm{T}}$ is the eigenvector. After normalization of the eigenvector $\boldsymbol{\upsilon}$, we get

$$\boldsymbol{\upsilon}^k \cdot \boldsymbol{\upsilon}^k = \langle\boldsymbol{\upsilon}^k, \boldsymbol{\upsilon}^k\rangle = 1, \quad k = 1, 2, \ldots, M \tag{7}$$

The principal components (scores) $\mathbf{t}_k$ $k = 1, 2, \ldots, M)$ can be expressed as

$$\mathbf{t}_k = \langle\boldsymbol{\upsilon}^k, \Phi(\mathbf{x})\rangle = \sum_{i=1}^{M}\alpha_i^k\mathrm{K}(\mathbf{x}, \mathbf{x}_i) \tag{8}$$

3.2. Base classifiers

In this paper, KPCA and SVM were combined to generate required number of base classifiers. Support vector machines are formulated based on the framework of statistical learning theory. They involve minimization of structural risk. As a two-class classification problem, SVM could separate the datasets by searching for an optimal separating hyperplane between them [31,32].

Given that the sample is repressed as $\mathbf{x}_i \in R^N$, $i = 1, 2, \ldots, M$ and each sample belongs to a class $y_i \in \{-1, 1\}$. For linear classification we can identify two classes by an optimal separating hyperplane

$$\mathbf{w}^{\mathrm{T}}\mathbf{x} + b = 0 \tag{9}$$

We can obtain the optimal values for $\mathbf{W}$ and b by solving a constrained convex quadratic programming problem, using

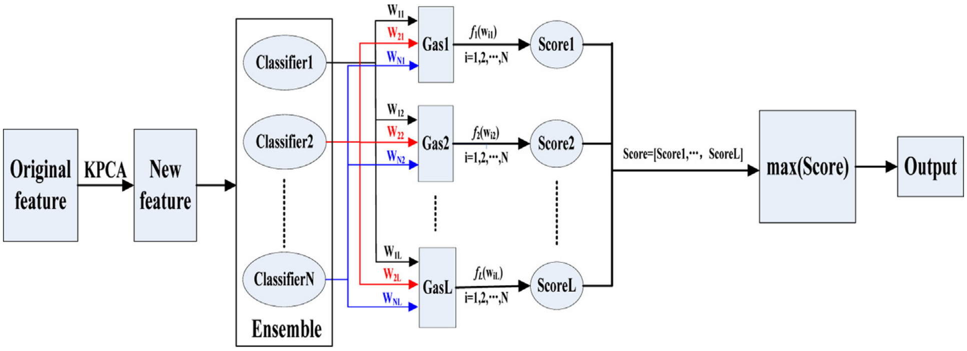

Fig. 3. The architecture of ISVMEN.

Lagrange multipliers $\alpha_i \quad (i = 1, 2, \ldots, M)$. The decision function can be expressed as

$$f(x) = \operatorname{sgn}\left(\sum_{i=1}^{M} \alpha_i y_i \mathrm{K}(\mathbf{x}_i, \mathbf{x}) + b\right) \tag{10}$$

where those $\mathbf{X}_i$ with nonzero α_i are the “support vectors”.

In most real-life applications linear separation of data sets cannot be achieved successfully. To overcome this, an alternative method is proposed. It works by first projecting original data to a higher dimensional space, and then performing subsequent analysis in the new space. If the mapping function is $\varphi(\mathbf{x})$, then the optimal separating hyperplane function can be written as:

$$f(\mathbf{x}) = \mathbf{w} \cdot \varphi(\mathbf{x}) + b \tag{11}$$

We can get a more generalized form as

$$f(x) = \sum_{i=1}^{M} \alpha_i \cdot y_i \cdot (\varphi(\mathbf{x}_i) \cdot \varphi(\mathbf{x})) + b \tag{12}$$

In a high dimensional space, determination of $\langle \varphi(\mathbf{x}_i), \varphi(\mathbf{y}_i) \rangle$ involves high computational cost. Fortunately, it turns out that this inner product can be evaluated using an appropriate kernel such that: $\mathrm{K}(\mathbf{x}_i, \mathbf{x}) = \langle \varphi(\mathbf{x}_i), \varphi(\mathbf{x}) \rangle$. In this paper, we use Gaussian RBF kernel which is defined as follows

$$\mathrm{K}(\mathbf{x}_i, \mathbf{x}_j) = \exp\left(\frac{-\left\|\mathbf{x}_i - \mathbf{x}_j\right\|^2}{\sigma^2}\right) \tag{13}$$

where σ^2 is the kernel parameter which determines the bandwidth of RBF. The decision function can by expressed as

$$f(x) = \operatorname{sgn}\left(\sum_{i=1}^{M} (\alpha_i \mathrm{K}(\mathbf{x}_i, \mathbf{x}) + b\right) \tag{14}$$

where α_i and b are the optimal decision parameters.

3.3. Multi-class ISVMEN in discrimination of indoor air contaminants

An effective and promising alternative to build a classifier is to train many classifiers and combine their decisions, which is called ensemble classifier system (ECS) and is currently among some of the hot research areas. In the course of building ECS, there are several important issues that could be grouped into three main problems [33]: (1) selection of the topology of the ECS; (2) design of base classifiers; (3) design of fusion (ensemble) method.

Usually, parallel topology is the most common topology because this structure has good methodological background [34]. The selection of base classifiers is also important. Indeed, ensemble learning will improve the generalization and prediction performance of individual classifiers only if these classifiers are accurate and diverse enough in their predictions. The fusion of base classifiers is the most important idea is classifiers ensemble. In this step, both the diversity among base classifiers and their accuracy are exploited to provide optimal combination (or fusion). Therefore, the choice of a collective decision making method is of paramount importance. There are two main fusion methods: majority voting and weight assignment. Majority voting utilizes the concept of democratic decision making. It states that the class label with more than half of base classifiers prediction is accepted as true [35]. The other approach is based on discriminant analysis. The main form of discrimination is a posterior probability typically associated with probabilistic pattern recognition model.

In this work, we focus on the issues of designing effective base classifiers and fusion method. A classifiers ensemble approach called as ISVMEN based on KPCA and SVM for classification of multiple indoor air pollutants using an E-nose was proposed. The architecture of ISVMEN model is depicted in Fig. 3.

In the proposed ISVMEN model, KPCA is first used for obtaining important nonlinear features from original data. KPCA results showed that the first 16 principal components accounted for 95% cumulative variance in original data. Thus, using these principal components, useful information from data sets of six kinds of air pollutants was extracted, which constitute the new features. Then, these new features were used as inputs of SVM multi-class classifier. In this paper, five base classifiers based SVM were designed. Finally, a weighted voting fusion was used to combine the classifiers. The implementation process of ISVMEN can be illustrated as follows.

Suppose that there are totally n training samples with p variables. Suppose also that the number of training samples for HCHO (formaldehyde), C_6H_6 (benzene), C_7H_8 (toluene), CO (carbon monoxide), NH_3 (ammonia), and NO_2 (nitrogen dioxide) is n_1, n_2, n_3, n_4, n_5 respectively. Thus, the initial training dataset can be represented by an $n \times p$ matrix as follows:

$$\mathbf{X}_{\text{old}} = \{\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3, \mathbf{X}_{4,}, \mathbf{X}_5, \mathbf{X}_6\} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix} \tag{15}$$

where the number of columns $p = 6$ and it represents the dimension of the original data; the number n of rows represents the total number of training samples; $\mathbf{X}_i \quad (i = 1, 2, \ldots, 6)$ represents the training data of the ith kind of gas.

It is worth noting that KPCA makes use of kernel trick to project original data $\mathbf{X}_{\text{old}}$ onto a higher dimensional feature space and gets the principal components as new features $\mathbf{X}_{\text{new}}$ by principal

Table 2
Results of KPCA.

Principle components	Accumulating contribution rate		
	Eigenvalue	Variance (%)	Cumulative variance (%)
PC1	68.07	29.95	29.95
PC2	53.78	23.66	53.61
PC3	22.23	9.78	63.40
PC4	16.99	7.47	70.88
PC5	13.33	5.86	76.75
PC6	11.12	4.89	81.64
PC7	7.60	3.34	84.99
PC8	4.66	2.05	87.04
PC9	3.75	1.65	88.69
PC10	2.89	1.27	89.97
PC11	2.87	1.26	91.23
PC12	2.41	1.06	92.29
PC13	2.23	0.98	93.28
PC14	1.92	0.84	94.13
PC15	1.66	0.73	94.85
PC16	1.39	0.61	95.47
PC17	0.98	0.43	95.90

component analysis. After KPCA, we can obtain more linearly separable features thereby getting better performance especially with limited training samples. Thus, the new training data matrix $\mathbf{X}_{\text{new}}$ can be expressed as

$$\mathbf{X}_{\text{new}} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1q} \\ x_{21} & x_{22} & \cdots & x_{2q} \\ \vdots & \vdots & & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nq} \end{bmatrix} \tag{16}$$

where q, the number of columns, depends on the number of principal components; the best value of q is obtained on the basis of cumulative variance (see Table 2) and prediction accuracy (see Fig. 2). Considering these two indicators, the best value of q was found to be 17.

The training goal is defined as

$$\mathbf{Y}_{\text{goal}} = [1, 2, 3, 4, 5, 6]^{\text{T}} \tag{17}$$

where 1, 2, 3, 4, 5 and 6 stands for the labels of HCHO, C_6H_6, C_7H_8, CO, NH_3, and NO_2, respectively.

Using these new features as input instances, a base classifier Ψ can classify any novel input instance as one of the predefined class labels $\mathbf{Y}_{\text{goal}}$. The base classifier Ψ is defined as [20]:

$$\Psi : \mathbf{X}_{\text{new}} \to \mathbf{Y}_{\text{goal}} \tag{18}$$

We use five classifiers $\Psi_1, \Psi_2, \Psi_3, \Psi_4, \Psi_5$. Each classifier as an expert makes decision according to its prediction accuracy on test dataset. The output from five base classifiers can be expressed as:

$$\hat{\mathbf{\Psi}} = [\Psi_1(\mathbf{x}), \Psi_2(\mathbf{x}), \ldots, \Psi_5(\mathbf{x})]^{\text{T}} \tag{19}$$

where $\Psi_i(\mathbf{x}) \in \mathbf{Y}_{\text{goal}},\quad i = 1, 2, \ldots, 5$ means the output of the ith base classifier on instance $\mathbf{x}$. To integrate decisions from base classifiers, we need to express the value $\Psi_i(\mathbf{x}) \in \mathbf{Y}_{\text{goal}},\quad i = 1, 2, \ldots, 5$ using binary encoding. The binary encoding method is shown as follows.If the output of the ith $(i = 1,\ 2, \ldots, 5)$ base classifier is HCHO namely $\Psi_i(\mathbf{x}) = 1$, then encoding it by $\mathbf{\Psi}_{\text{code}i}(\mathbf{x}) = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}^{\text{T}}$;

Similarly,

if $\Psi_i(\mathbf{x}) = 2$, then encoding it by $\mathbf{\Psi}_{\text{code}i}(\mathbf{x}) = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 \end{bmatrix}^{\text{T}}$;
if $\Psi_i(\mathbf{x}) = 3$, then encoding it by $\mathbf{\Psi}_{\text{code}i}(\mathbf{x}) = \begin{bmatrix} 0 & 0 & 1 & 0 & 0 & 0 \end{bmatrix}^{\text{T}}$;
if $\Psi_i(\mathbf{x}) = 4$, then encoding it by $\mathbf{\Psi}_{\text{code}i}(\mathbf{x}) = \begin{bmatrix} 0 & 0 & 0 & 1 & 0 & 0 \end{bmatrix}^{\text{T}}$;
if $\Psi_i(\mathbf{x}) = 5$, then encoding it by $\mathbf{\Psi}_{\text{code}i}(\mathbf{x}) = \begin{bmatrix} 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix}^{\text{T}}$;
if $\Psi_i(\mathbf{x}) = 6$, then encoding it by $\mathbf{\Psi}_{\text{code}i}(\mathbf{x}) = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}^{\text{T}}$;

Then we have

$$\hat{\mathbf{\Psi}}_{\text{code}} = [\mathbf{\Psi}_{\text{code}1}(\mathbf{x}), \mathbf{\Psi}_{\text{code}2}(\mathbf{x}), \ldots, \mathbf{\Psi}_{\text{code}5}(\mathbf{x})]^{\text{T}} = \begin{bmatrix} \hat{\Psi}_{\text{code}11} & \cdots & \hat{\Psi}_{\text{code}16} \\ \vdots & & \vdots \\ \hat{\Psi}_{\text{code}51} & \cdots & \hat{\Psi}_{\text{code}56} \end{bmatrix} \tag{20}$$

where $\hat{\mathbf{\Psi}}_{\text{code}}$ is a matrix with 5×6 and the element $\hat{\Psi}_{\text{code}ij}\quad (i = 1, 2, \ldots, 5;\quad j = 1, 2, \ldots, 6)$ is the result of the ith classifier on the jth gas after binary encoding; specifically, each row represents a classifier while each column represents a target gas.

Suppose that the total weight matrix of five base classifiers is expressed by

$$\mathbf{W} = [\mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_3, \mathbf{w}_4, \mathbf{w}_5]^{\text{T}} = \begin{vmatrix} w_{11} & w_{12} & w_{13} & w_{14} & w_{15} & w_{16} \\ w_{21} & w_{22} & w_{23} & w_{24} & w_{25} & w_{26} \\ w_{31} & w_{32} & w_{33} & w_{34} & w_{35} & w_{36} \\ w_{41} & w_{42} & w_{43} & w_{44} & w_{45} & w_{46} \\ w_{51} & w_{52} & w_{53} & w_{54} & w_{55} & w_{56} \end{vmatrix} \tag{21}$$

where $w_{ij}\quad (i = 1, 2, \ldots, 5;\quad j = 1, 2, \ldots, 6)$ means the weight of the ith base classifier on the jth gas. The entries in each row represent the weights assigned to each base classifier. It is worth mentioning that the same classifier may have different weights for the six gases. The weights of each base classifier were computed as follows [26]

$$w_{ij} = \frac{\log(p_{ij}/(1 - p_{ij}))}{\sum_{i=1}^{5} \log(p_{ij}/(1 - p_{ij}))},\quad i = 1, 2, \ldots, 5;\quad j = 1, 2, \ldots, 6 \tag{22}$$

where p_{ij} is the accuracy of the ith classifier on the jth gas and it was obtained from test dataset. It has been proved that in literature [36] the ensemble classifier system's classification accuracy could reach the maximum and take full advantage of the prior information of base classifiers if one uses Eq. (22) to calculate the weights.

The classifiers ensemble system makes final decision based on Eqs. (23) and (24)

$$score_j = f_j(w_{ij}, \hat{\Psi}_{\text{code}ij}) = \sum_{i=1}^{5} w_{ij} \cdot \hat{\Psi}_{\text{code}ij},\quad j = 1, 2, \ldots, 6 \tag{23}$$

where $score_j$ is the score of the jth gas. That is, each gas has its own score, and the label of the gas can be represented by the maximum score

$$Gas\_label = \max(score_1, score_2, score_3, score_4, score_5, score_6) \tag{24}$$

4. Results and discussion

All the computations are carried out in MATLAB R2011b software. In order to make an impartial comparison between base classifiers and the proposed ensemble classifier, we evaluated in terms of the classification accuracy on validation dataset (see Table 1) which was mainly used to validate the performance of ensemble classifier. The specific process and approach of obtaining the training set, test set, validation set have been described in Section 2.3. The training set was used to generate base classifiers and we assess the classification accuracy of base classifiers

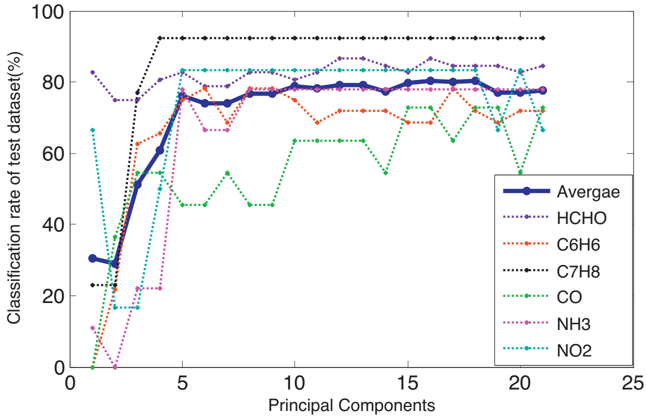

Fig. 4. Classification rate of test dataset in terms of number of principal components. (For interpretation of the references to color in the text, the reader is referred to the web version of this article.)

Table 3
Comparison of classification accuracy on test dataset for four feature extraction methods.

Class	Classification accuracy (%) on test set			
	Original	PCA	ICA	KPCA
HCHO	90.3	94.2	66.5	92.3
C_6H_6	84.8	75.7	59.5	63.6
C_7H_8	100.0	92.3	51.2	84.6
CO	58.3	83.3	28.8	100.0
NH_3	60.0	40.0	63.3	100.0
NO_2	66.6	83.3	40.0	50.0
Average	76.8	78.1	51.3	**81.7**

using homologous testing set. Based on base classifiers' classification accuracy, we can obtain the weights for each base classifier using Eq. (22). From Table 1, it is easy to find that the dataset for six kinds of contaminants is quite imbalanced. The total number of HCHO and C_6H_6 are 260,164 respectively; however, the total number of the remaining four gases is far less than them. Especially, NO_2 has only 30 samples. In this case, a single classifier could hardly obtain good classification results for all gases. That is, the class with fewer samples is easier to be classified as the class with more samples.

KPCA results of the training samples are presented in Table 2. Table 2 shows the eigenvalues and the cumulative variance of the first 17 principal components. From this table, we can see that the variance contribution of the first 16 components reaches 95.47% which means that the sample information can be represented fully by using the first 16 principal components. In the implementation of KPCA, there are two major issues to be solved: one is to search for the optimal number of principal components, and the other is to find the best value of kernel parameter σ. In the beginning, the first principal component was used as input of SVM. Then the number of inputs (i.e. the number of principal components) of SVM increased gradually, up to 22 inputs. The reason why we stop at 22 principal components is that the first 16 principal components have explained more than 95% of the total variance. Then it is unnecessary to increase the principal components continually. Fig. 4 illustrates the variation trend of classification accuracy versus the number of principal components. From this figure, one can notice that the best average accuracy was obtained at point 17 and point 19, as indicated by the blue line. However, considering the dimension aspect, we finally adopted 17 principal components. In addition, we have also compared KPCA with two linear feature extraction methods namely PCA, and ICA. Experimental results in Table 3 showed the superiority of KPCA over the linear feature extraction methods. In Table 3, KPCA obtained the best classification accuracy of 81.7% with bold type, however, the mean accuracy of raw feature, PCA and ICA are only 76.8%, 78.1%, 51.3% respectively. Besides, for a single gas, KPCA had a good classification accuracy of 100%, 100%, and 92.3% for CO, NH_3 and HCHO respectively. By taking into account the mean recognition rate and the single gas recognition rate, finally, we used KPCA as the feature extraction method to improve the base classifier's accuracy.

Table 5
Comparison of classification accuracy for six models using validation dataset.

Class	Classification accuracy (%) of validation set						
	SVM1	SVM2	SVM3	SVM4	SVM5	MV	ISVMEN
HCHO	84.6	88.4	86.5	92.3	92.3	92.3	90.3
C_6H_6	78.7	81.8	81.8	78.7	63.6	81.8	81.8
C_7H_8	100.0	100.0	100.0	84.6	84.6	100.0	100.0
CO	83.3	100.0	91.6	91.6	100.0	83.3	100.0
NH_3	60.0	100.0	90.0	90.0	100.0	100.0	100.0
NO_2	83.3	50.0	66.6	50.0	50.0	83.3	83.3
Average	81.6	86.7	86.1	81.2	81.7	90.1	92.5

Table 4 shows the classification accuracy of five base classifiers for training and test samples. The five base classifiers were labeled SVM1, SVM2, SVM3, SVM4, and SVM5, respectively. From this table, we can see that the training results are desirable and all the mean accuracy exceeded 93%, however, the test results are not uniformly good. For SVM1 and SVM5, SVM1 has a good accuracy of 100.0% on toluene but a poor accuracy of 77.7% on ammonia; the situation for SVM5 is reversed. Besides, for the five base classifiers, the best classification accuracy on NO_2 and C_6H_6 are only 66% and 75% respectively; the worst classification accuracy on NO_2 and C_6H_6 are low to 50% and 59% respectively. Therefore, the single classifiers are not uniformly good for all gases. More specifically, the classification precision and generalization ability of single classifier can hardly meet our requirement.

The performance of the proposed multiple classifiers system and that of the base classifiers were evaluated on a validation dataset which has never been used (see Table 5). In addition, the

Table 4
Classification accuracy of five base classifiers for training and testing samples.

Class	Classification accuracy (%)									
	Train set					Test set				
	SVM1	SVM2	SVM3	SVM4	SVM5	SVM1	SVM2	SVM3	SVM4	SVM5
HCHO	94.8	98.0	92.9	96.1	94.2	73.0	82.6	84.6	76.9	86.5
C_6H_6	95.9	95.9	88.8	88.8	93.9	65.6	75.0	75.0	65.6	59.3
C_7H_8	100.0	100.0	100.0	97.5	100.0	100.0	92.3	100.0	100.0	84.6
CO	100.0	97.1	97.1	97.1	100.0	81.8	90.9	81.8	100.0	100.0
NH_3	93.1	100.0	100.0	96.5	100.0	77.7	100.0	88.8	88.8	100.0
NO_2	94.4	94.4	94.4	83.3	94.4	66.6	66.6	66.6	50.0	50.0
Average	96.3	97.6	95.5	93.2	97.1	77.4	84.5	82.8	80.2	78.5

Table 6
Multi-class classification results on validation set using ISVMEN model.

Class	*N*	Classified as					
		HCHO	C_6H_6	C_7H_8	CO	NH_3	NO_2
HCHO	52	**47**	5	0	0	0	0
C_6H_6	33	5	**27**	0	1	0	0
C_7H_8	13	0	0	**13**	0	0	0
CO	12	0	0	0	**12**	0	0
NH_3	10	0	0	0	0	**10**	0
NO_2	6	1	0	0	0	0	**5**

performance of another fusion mechanism, majority voting, was evaluated. Majority voting (labeled as MV, see Table 5) uses the concept of democratic decision making wherein all base classifiers have the same weight which is the reciprocal of the number of base classifiers.

From Table 5, it can be seen that the classification accuracy of the proposed method (ISVMEN) is higher than that of any of the base classifiers. Of course one base classifier may have good performance on several gases. On the other hand, the proposed method (ISVMEN) can obtain the best discrimination accuracy for almost all gases. For C_7H_8, CO and NH_3, almost 100% classification accuracy was obtained using our method. Due to limited number of samples, four out of five base classifiers had classification accuracy less than 70% for NO_2. However, ISVMEN model the proposed method can overcome sample unbalanced problem and classify each gas more accurately. Also, the average classification accuracy has been improved from less than 86% to 92.58%. At the same time, compared to MV fusion method (majority voting), the proposed fusion method can obtain better result, and average recognition accuracy has achieved 92.58% which are higher than 90.1% obtained using MV with an increment of two percent in classification accuracy. In conclusion, the ensemble classifier proposed in current work is superior to base classifiers for both classification accuracy and generation ability.

Table 6 presents the classification results of validation samples using ISVMEN model. The digits with bold type in diagonal line denote the number of correctly classified samples, and others mean the number of misclassified samples. It is clear to find that almost all the gases can be classified correctly.

According to the front analysis, it is easy to reach the conclusion that the proposed method ISVMEN is better favorably with other methods. Two main reasons can explain it: (1) using nonlinear KPCA to extract features and improve the precision of base classifiers. In our previous publications [9], it can be seen that the six gases are linearly inseparable, then linear feature extraction methods (i.e. PCA, ICA) are unable to obtain helpful nonlinear feature. In Table 3, it is easy to find that the classification accuracy of base classifier SVM has been greatly improved after KPCA. (2) The use of improved weighted fusion approach results in highly diverse base classifiers. The prediction accuracy of each base classifier depends on the gases to be predicted. For example, for classifier 1 which refers to SVM1 in Table 4, the prediction accuracy on C_7H_8 is very high, but it has a poor accuracy on C_6H_6. If uniform weights are used for all gases just as MV wherein all base classifiers have the same weight which is the reciprocal of the number of base classifiers, then diversity among base classifiers will not be fully exploited. As to the common fusion method MV, all the base classifiers would be assigned an equal weight, and then one cannot take full advantage of their diversity. Therefore, we adopt improved weighted approach in which weights assignment is done based on the predictive accuracy of each base classifier on each gas. After the weights for each base classifier have been determined, we used Eq. (23) to combine the decisions from all base classifiers thereby obtaining final score for each gas. In decision level, the gas with the highest score was considered as the winner.

In spite of the good results, we can continue our research in the following aspect: in order to improve the base classifier's accuracy sequentially the optimization algorithm such as GA (Genetic Algorithm) can be used to obtain both the optimal kernel parameter and the best number of principal components in terms of KPCA. So, it leaves some space for further study in the subsequent development of theory and practice.

5. Conclusion

A novel multi-class recognition approach ISVMEN was investigated in this paper for classification of multiple indoor air contaminants. It is based on multiple classifier systems in which KPCA and SVM were combined to build the base classifiers, and an effective fusion strategy was presented to integrate the decisions from the base classifiers. The purpose of this work is to improve the prediction accuracy and robustness of the pattern recognition scheme in E-nose. The performance of the proposed method was compared with that of base classifiers, and also that of standard majority voting. Experimental results show that, average recognition accuracy has achieved 92.58% which is higher than 86%, the best result among the base classifiers. Furthermore, the proposed fusion method can obtain better result compared to MV fusion method (majority voting) which has an average recognition accuracy of 90.1%. Comparatively, the result of this work demonstrates that the proposed ISVMEN model can achieve a better performance both in recognition accuracy and generalization compared with that of base classifiers, and also that of standard majority voting.

Acknowledgements

We would like to express our sincere appreciation to the anonymous reviewers for their insightful comments, which have greatly improved the quality of the paper.

This work was supported by the Key Science and Technology Research Programs (Nos. CSTC2010AB2002 and CSTC2009BA2021). This work was also funded by Hong Kong Scholar Program (No. XJ2013044) and New Academic Researcher Award for Doctoral Candidates granted by Ministry of Education in China.

References

[1] U. Siripatrawan, Rapid differentiation between *E. coli* and *Salmonella typhimurium* using metal oxide sensors integrated with pattern recognition, Sensors and Actuators B 133 (2008) 414–419.

[2] L. Zhang, F. Tian, C. Kadri, G. Pei, H. Li, L. Pan, Gases concentration estimation using heuristics and bio-inspired optimization models for experimental chemical electronic nose, Sensors and Actuators B 160 (2011) 760–770.

[3] L.I. Kuncheval, Combining pattern classifiers: methods and algorithms, IEEE Transactions on Neural Networks 18 (2007) 964.

[4] Y. Su, S. Shan, X. Chen, W. Gao, Hierarchical ensemble of global and local classifiers for face recognition, IEEE Transactions on Image Processing 18 (2009) 1885–1896.

[5] D. Tao, X. Tang, et al., Asymmetric bagging and random subspace for support vector machines-based relevance feedback in image retrieval, IEEE Transactions on Pattern Analysis and Machine Intelligence 28 (2006) 1088–1099.

[6] A. Bermak, D. Martinez, A compact 3D VLSI classifier using bagging threshold network ensembles, IEEE Transactions on Neural Networks 14 (5) (2003) 1097–1109.

[7] M. Shi, S. Brahim-Belhouari, A. Bermak, Quantization errors in committee machine for gas sensor application, IEEE International Symposium on Circuits and Systems 3 (2005) 1911–1914.

[8] M. Shi, S. Brahim-Belhouari, A. Bermark, D. Martinez, Committee machine for odor discrimination in gas sensor array, in: Proceedings of the 11th International Symposium on Olfaction and Electronic Nose (ISOEN), 2005, pp. 74–76.

[9] S. Bermejo, J. Cabestany, Ensemble learning for chemical sensor arrays, Neural Processing Letters 19 (2004) 25–35.

[10] D.Q. Gao, M.M. Chen, J. Yan, Simultaneous estimation of classes and concentrations of odors by an electronic nose using combinative and modular multilayer perceptrons, Sensors and Actuators B 107 (2005) 773–781.

[11] D.Q. Gao, W. Chen, Simultaneous estimation of odor classes and concentrations using an electronic nose with function approximation model ensembles, Sensors and Actuators B 120 (2007) 584–594.

[12] V. Hirayama, F.J. Ramirez-Fernandez, W.J. Salcedo, Committee machine for LPG calorific power classification, Sensors and Actuators B 116 (2006) 62–65.

[13] E. Bona, R. Silva, D. Borsato, D.G. Bassoli, Neural network for instant coffee classification through an electronic nose, International Journal of Food Engineering 7 (2011), http://dx.doi.org/10.2202/1556-3785.2002.

[14] A. Vergara, S. Vembu, T. Ayhan, A.R. Margaret, L.H. Margie, H. Ramón, Chemical gas sensor drift compensation using classifier ensembles, Sensors and Actuators B 166–167 (2012) 320–329.

[15] A. Amini, M.A. Bagheri, G. Montazer, Improving gas identification accuracy of a temperature-modulated gas sensor using an ensemble of classifiers, Sensors and Actuators B 187 (2013) 241–246.

[16] B. Scholkopf, A. Smola, K.R. Muller, Nonlinear component analysis as a Kernel eigenvalue problem, Neural Computation 10 (1998) 1299–1319.

[17] Z.L. Sun, D.S. Huang, Y.M. Cheun, Extracting nonlinear features for multispectral images by FCMC and KPCA, Digital Signal Processing 15 (2005) 331–346.

[18] L.J. Cao, K.S. Chua, W.K. Chong, et al., A comparison of PCA, KPCA and ICA for dimensionality reduction in support vector machine, Neurocomputing 55 (2003) 321–336.

[19] L. Zhang, F.C. Tian, H. Nie, et al., Classification of multiple indoor air contaminants by an electronic nose and a hybrid support vector machine, Sensors and Actuators B 174 (2012) 114–125.

[20] L. Zhang, F.C. Tian, S. Liu, et al., Chaos based neural network optimization for concentration estimation of indoor air contaminants, Sensors and Actuators A 189 (2013) 161–167.

[21] C. Kadri, F.C. Tian, L. Zhang, et al., Neural network ensembles for online gas concentration estimation using an electronic nose, International Journal of Computer Science Issue 10 (2013) 129–135.

[22] F.C. Tian, H.J. Li, L. Zhang, et al., A denoising method based on PCA and ICA in electronic nose for gases quantification, Journal of Computational Information Systems 8 (2012) 5005–5015.

[23] L. Zhang, F. Tian, C. Kadri, B. Xiao, H. Li, L. Pan, H. Zhou, On-line sensor calibration transfer among electronic nose instruments for monitoring volatile organic chemicals in indoor air quality, Sensors and Actuators B 160 (2011) 899–909.

[24] I.T. Jolliffe, Principal Component Analysis, second ed., Springer-Verlag, New York, 2002.

[25] A. Hyvärinen, E. Oja, Independent component analysis: algorithms and applications, Neural Networks 13 (2000) 411–430.

[26] S. Mika, B. Scholkopf, A.J. Smola, Kernel PCA and de-nosing in feature spaces, Advances in Neural Information Processing Systems 11 (1999) 536–542.

[27] J.H. Cho, J.M. Lee, S.W. Choi, D. Lee, I.B. Lee, Fault identification for process monitoring using kernel principal component analysis, Chemical Engineering Science 60 (1) (2005) 279–288.

[28] S.W. Choi, C. Lee, J.M. Lee, J.H. Park, I.B. Lee, Fault detection and identification of nonlinear process based on kernel PCA, Chemometrics and Intelligent Laboratory Systems 75 (1) (2005) 55–67.

[29] J.D. Shao, G. Rong, J.M. Lee, Learning a data-dependent kernel function for KPCA-based nonlinear process monitoring, Chemical Engineering Research and Design 87 (2009) 1471–1480.

[30] Y.P. Zheng, L.P. Zhang, Fault diagnosis of wet flue gas desulphurization system based on KPCA, in: The 19th International Conference on Industrial Engineering and Engineering Management, 2013, pp. 279–288.

[31] E. Gumus, N. Kilic, A. Sertbas, et al., Evaluation of face recognition techniques using PCA, wavelets and SVM, Expert Systems With Applications 37 (2010) 6404–6408.

[32] N. Cristianini, J. Taylor, An Introduction to Support Vector Machines and Other Kernel-based Learning Methods, Cambridge University Press, Cambridge, United Kingdom, 2000.

[33] M. Wozniak, M. Zmyslony, Designing combining classifier with trained fuser–analytical and experimental, Neural Network Word 20 (2010) 807–978.

[34] L.I. Kuncheva, Combining pattern classifiers: methods and algorithms, IEEE Transactions on Neural Networks 18 (2004) 964.

[35] A. Szczurek, B. Krawczyk, M. Maciejewska, et al., VOCs classification based on the committee of classifiers coupled with single sensor signals, Chemometrics and Intelligent Laboratory Systems 125 (2013) 1–166.

[36] L.I. Kuncheva, C.J. Whitaker, C.A. Shipp, et al., Limits on the majority vote accuracy in classifier fusion, Pattern Analysis and Applications 6 (2003) 22–31.

Biographies

Lijun Dang received her Bachelor degree in School of Electronic and Information Engineering in 2011 from the Dalian University of Technology, China; she is now pursuing her MS degree in circuits and system. Her research interests include circuits and system design in electronic nose technology.

Fengchun Tian received Ph.D. degree in 1997 in electrical engineering from Chongqing University. He is currently a professor with the College of Communication Engineering of Chongqing University. His research interests include Electronic nose technology, artificial olfactory systems, pattern recognition, chemical sensors, signal/image processing, wavelet, and computational intelligence. In 2006 and 2007, he was recognized as a part-time Professor of GUELPH University, Canada.

Lei Zhang received his Ph.D. degree in 2013 in Circuits and Systems from Chongqing University. He was selected as a Hong Kong scholar in China in 2013. In the same year, he obtained the Youth Innovation Award of Academician. He was also honored by New Academic Researcher Award for Doctoral Candidates granted by Ministry of Education in China in 2012. Now he works as postdoctoral fellow in the Hong Kong Polytechnic University. His current research interests include machine learning, pattern recognition, artificial olfactory system, and nonlinear signal processing in Electronic nose.

Chaibou Kadri received his Ph.D. degree in Circuits and Systems in 2013 from Chongqing University of China. His research interests include signal processing for gas sensors array instruments, and machine learning.

Xin Yin received his Bachelor degree in Communication Engineering in 2012 from the Chongqing University, China; he is not pursuing his MS degree in circuits and system. His research interests include electronic nose and circuit design.

Xiongwei Peng received his Bachelor degree in Communication Engineering in 2012 from the Chongqing University, China; he is now pursuing his MS degree in circuits and system. His research interests include artificial neural network and optimizations.

Shouqiong Liu is now a senior engineer of Academy of Metrology and Quality Inspection, Chongqing. Her research interest was mainly analytical chemistry.