A novel pattern mismatch based interference elimination technique in E-nose

Fengchun Tian, Zhifang Liang*, Lei Zhang, Yan Liu, Zhenzhen Zhao

College of Communication Engineering, Chongqing University, 174 ShaZheng Street, ShaPingBa District, Chongqing 400044, China

ABSTRACT

Metal oxide semiconductor (MOS) sensor array with cross-sensitivity to target gases is often used in electronic nose (e-nose) for monitoring indoor air quality. However, MOS sensors have their own defects of high susceptibility to some interferences which would seriously impact on the detection of target gases. Therefore it is urgent to solve the problem of interferences elimination as e-nose composed of MOS sensors cannot be used when there are interferences. A closely related method tends to discriminate the interference gases and target gases, and it depends on the type of interference gases. However, there are numerous interferences in real-world application scenario, which is impossible to be sampled in laboratory experiments. Considering that target gases detected by an e-nose can be fixed as invariant information, a novel and effective *Pattern Mismatch based Interference Elimination* (PMIE) method is proposed in this paper. It contains two parts: interference discrimination (i.e. pattern mismatch) and correction (i.e. interference elimination). Specifically, the principle of interference discrimination is whether a new pattern violates the rules established on the invariant target gases information (i.e. the case of interference gas appearing) or not (i.e. the case of only target gas appearing). If the current pattern of the sensor array is of interference, orthogonal signal correction algorithm (OSC) is used for interference correction. Experimental results prove that the proposed PMIE method is significantly effective for interference elimination in e-nose.

© 2016 Elsevier B.V. All rights reserved.

1. Introduction

Electronic nose (e-nose), as an artificial nose, is a technology in modeling of biological olfactory system. In recent years, e-nose techniques have been widely used in many areas, such as environment monitoring, food quality evaluation, and industrial control etc. In this paper, a developed e-nose which consists of metal oxide semiconductor (MOS) gas sensors and some pattern recognition algorithms for indoor air quality monitoring is introduced. The performance of e-nose depends largely on the selected sensors, which should have good cross-sensitivity, selectivity, reliability and robustness [1].

Objectively, the cross-sensitivity of sensor array has both advantages and disadvantages. Specifically, the cross-sensitivity is beneficial for detecting many kinds of gases with a limited number of gas sensors. However, this also results in that the sensor array in an e-nose would produce significant responses to some interferences. Thus the interferences will have a serious impact on the

detection of the target gases. It is hopeful that the interferences can be removed by designing appropriate filter, however, it cannot be effectively separated from target responses by traditional filters because interference responses are almost entirely mixed with the target responses with similar frequencies. To solve the problem of interference elimination, the machine learning idea is utilized in this paper.

Generally, there are two kinds of interferences in an e-nose. One is from those non-target odors which do not belong to the target gases being detected. The other one is from environmental factors, such as temperature, humidity, etc. [2]. In conventional e-noses, the latter, i.e., temperature and humidity have been compensated by integrating corresponding sensors into the e-nose sensor array. Currently the methods for eliminating the impacts from environmental temperature and humidity are based on temperature and humidity compensation model [3–8]. A physical way, where SiO_2 is used as sensing material with a titanium thermistor being used to maintain a constant temperature, is used to compensate temperature and humidity influences [9]. Independent component analysis (ICA) is also used for reducing the environmental impacts [10], where the component with the maximum correlation coefficient with regard to the reference vector is removed as interference. However, the non-target odor interferences cannot be handled

Table 1
Target gas samples for models setting up.

Target gas	formaldehyde	toluene	benzene	NO_2	NH_3	CO
Number of Samples	504	264	288	152	116	232
Number of train/test samples	336/168	176/88	192/96	101/51	77/39	155/77

effectively so far, which seriously limits the real-time application of e-nose in real-world scenario, such as the mall with perfumes, the kitchen with smoke flavor, etc. since the MOS gas sensors are very sensitive to pungent odors like perfumes and smoke flavor.

The problem of non-target odors interference is urgent to be solved in e-nose, while the conventional methods that tend to solve this problem can be divided into two steps: interferences discrimination and interferences correction [11]. Briefly, the first step is to determine that whether the current e-nose response is from interference or not. If yes, the second step is performed to correct the sensor responses with interference elimination method. Therefore, a classifier that is used to discriminate the interference should be first set up by the limited types of interference samples, such that the learned classifier largely depends on the limited number of interferences and it may be only effective to the limited types of interference. However, there are so many possible kinds of interferences that cannot be collected in laboratory experiments. Therefore, conventional methods cannot discriminate all non-target odor interferences in real-world application scenarios. In the correction step, generally the correction of interference response depends on the previous target response, so that the e-nose cannot work in real time when interferences exist.

As shown in Ref. [12] for interferences elimination of e-nose in wound detection, it tends to remove the background interference according to the spatial correlation coefficients by performing wavelet transform on the collected odor samples between the infected and healthy mice. In Ref. [13], the sensor response is analyzed by independent component analysis (ICA) and the sensor response of healthy mice body odor measured independently is regarded as a reference vector of background interference, then the correlation coefficients between the components of ICA and the reference vector are calculated where the components with the biggest coefficient is regarded as background. It is known that, in this method, the reference vector is very important for determining the interference component. However, interferences are various in complex real-world scenario, such that it is difficult to obtain the so-called reference vector. Additionally, the ICA algorithm shows a good effect in eliminating environment interferences, but it does not fit the non-target gas interferences. One reason may be that the target gas and non-target gas have similar nature and influence on the array of sensors, and cannot be effectively separated from the mixed responses. The orthogonal signal correction algorithm (OSC) may remove the information that is orthogonal to target

Table 2
Training algorithm of the proposed PMIE method.

Algorithm 1. PMIE training algorithm

Input
　Dataset 1 ($\mathbf{X}_1$);
　Dataset 2 ($\mathbf{X}_2$);

Output
　Prediction function $f(\cdot)$;
　Threshold T.

Procedure
　1. Data preprocessing;
　2. Obtain the prediction function $f(\cdot)$ by training on $\mathbf{X}_1$;
　3. For $i = 1$ to N
　　Predict the fourth sensor response: $y = f(x_{\text{tem}}, x_{\text{hum}}, x_{s1}, x_{s2}, x_{s3})$ by using $\mathbf{X}_2$; Calculate the prediction error: $error_i = |y - x_{s4}|$
　end
　4. Calculate the misjudgment rates of the 100 thresholds;
　5. Get the threshold T with the smallest misjudgment rate.

---

signal, where the orthogonal information is recognized as irrelevant information, therefore, it has been used to correct signals and remove interference information in [14,15]. However, the problem of determining the reference vector still exists as ICA. Moreover, the interference in Ref. [15] is fixed in the whole detection. Various interferences in real-world environment that are independent of the target gases are considered in our proposed method.

From the analysis above, many appropriate methods are developed to compensate and eliminate environmental interferences. But the interferences of non-target gases are still troubling. In order to solve this problem, a novel Pattern Mismatch based Interference Elimination (PMIE) method consisting of discrimination and correction is proposed in this work.

The rest of the paper is organized as follows. The experimental setup and experimental data are introduced in Section 2. The PMIE method proposed in this paper is introduced in Section 3. Experimental results and discussion are presented in Section 4. Conclusion is given in Section 5.

2. Experiments

2.1. Experimental setup

The experimental platform and the designed e-nose system are described in Fig. 1. The e-nose system is composed of an array of
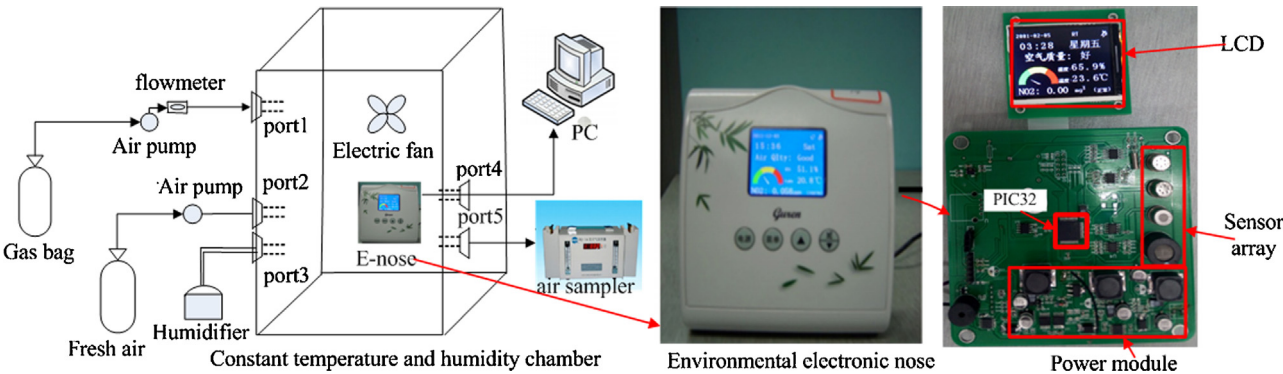

Fig. 1. Experimental platform and e-nose system designed for environmental monitoring.

sensors (TGS2602, TGS2620 and TGS2201 with two outputs A and B), a CPU (PIC32), a power module, a LCD display, etc. Besides, a module with two sensors for temperature and humidity is also used as the inputs of model for compensating the influence of environment factors. The analog-digital converter is used as interface between the sensors and PIC32 processor. In this paper, six target gases are monitored by this environmental e-nose, they are formaldehyde (HCHO), benzene (C_6H_6), toluene (C_7H_8), carbon monoxide (CO), ammonia (NH_3) and nitrogen dioxide (NO_2). The experimental platform consists of a programmable temperature and humidity chamber, a gas sampler, a personal computer, an air pump, a humidifier, etc. Specifically, the programmable temperature and humidity chamber can provide a stable experimental environment, the air sampler is used for sampling the gas within chamber and determining the concentration of the gas using spectrophotometer and gas chromatography (GC) analysis, the humidifier is for adjusting the humidity in the chamber, and the personal computer is used to collect and process the data from the e-nose system.

In the preparation stage, formaldehyde, benzene and toluene are liquid, they were volatilized in the gas collecting bags after their liquor are injected, respectively. While NH_3, CO, NO_2 are gases, so the standard gases were injected into the gas collection bags, directly. These target gases in the gas collection bags were diluted by pure nitrogen (N_2). In experiments, the quantity of gas was controlled by injection time through a flow-meter connected with the standard gas bag. The experimental temperature was set to 15, 20, 25, 30, 35 and 40 °C, and the relative humidity was set to 40%, 60% and 80%, respectively. The total time of each experiment was 10 min including 2 min for baseline and 8 min for steady state. For each sample, 300 sampling points were collected. After each experiment, the chamber was purged by clean air until the response of sensors returned to the baseline state.

2.2. Experimental data

Three datasets are prepared respectively, as follows.

2.2.1. Dataset 1
Dataset 1 shown in Table 1 is a multi-class sample set composed of six target gases. Each sample is the steady state point in the sensor response, that is, each sample is generated by an individual sensor response sequence. The experimental process is illustrated in Section 2.1. It is used to train a prediction model by dividing into training and test data. The training samples are selected from Dataset 1 by using KS algorithm [16] and the remaining examples
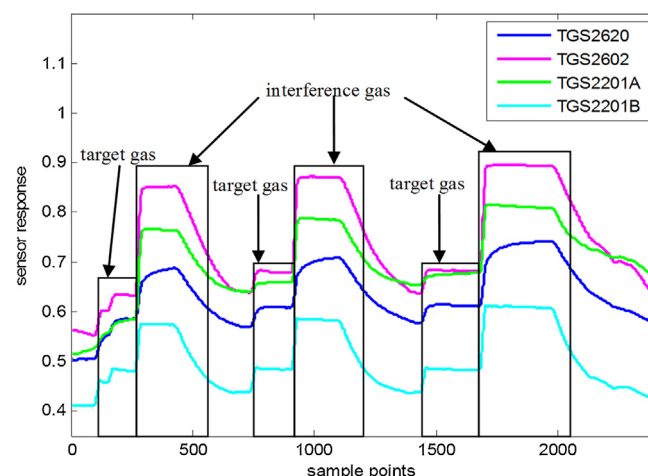

Fig. 2. Response curve of the four sensors of Dataset 3.

are used as test data. The proportion of training and testing data is approximately 2:1 for generalization.

2.2.2. Dataset 2
It is used to determine the threshold and test the performance of the prediction model, which includes target gas samples and interference gas samples. The experimental method for obtaining Dataset 2 is the same as that of Dataset 1. Note that the interference experiments and target experiments were carried out separately. The experimental process is illustrated in experimental setup. Here perfume and alcohol are used as the interferences. The number of target and interference gas samples is 1205 and 1195, respectively.

2.2.3. Dataset 3
It is a data set for verifying the performance of the PMIE model in real environment. It is obtained when the e-nose is exposed to the mixture of interferences and the target gas. The experimental process is shown as follows. First, the target gas is injected into the chamber, until it reaches the steady state. Then the interference gases are injected into the chamber. Third, repeat these two steps three times. The selected target gas is formaldehyde, and the interferences are perfume, alcohol and orange, respectively. In dataset 3, there are 2400 points in total, which simulates the case of interference in real E-nose application. All the data points are used to verify the performance of PMIE model in real environment. The curves of sensor responses are described in Fig. 2.
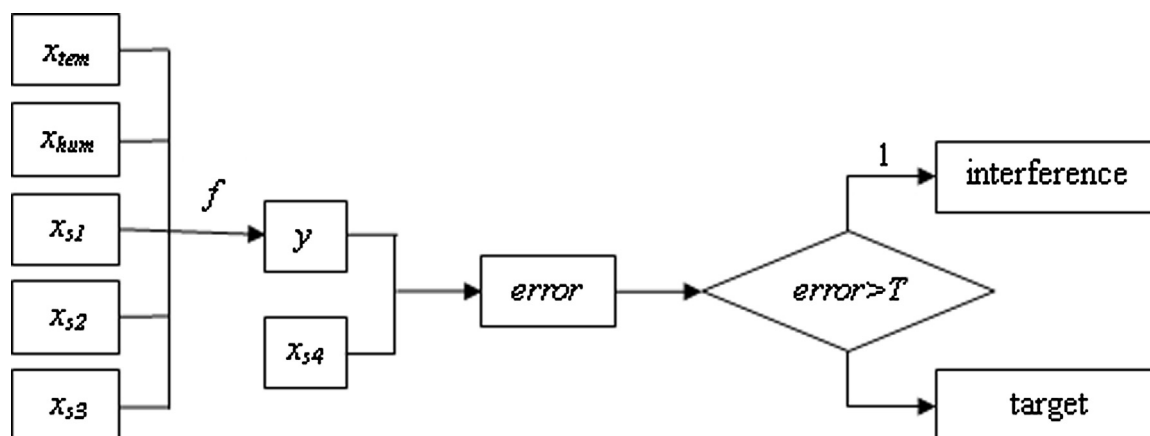

Fig. 3. Structure of PMIE for interference discrimination.

Table 3
Testing algorithm of the proposed PMIE method.

Algorithm 2. PMIE testing algorithm
Input
The prediction function $f(\cdot)$;
The threshold T;
Dataset 3 ($\mathbf{X}_3$).
Output
The interference *label*;
The interference eliminated sensor response $\mathbf{Z}_{cor}$.
Procedure
1. Data preprocessing;
2. Predict the fourth sensor response: $y = f(x_{tem}, x_{hum}, x_{s1}, x_{s2}, x_{s3})$ by using $\mathbf{X}_3$;
3. Calculate the prediction error: $error = |y - x_{s4}|$;
4. Predict the interference *label* based on pattern mismatch:
if $error > T$
label = 1;
else *label* = 0;
5. Correct the original signal $\mathbf{X}_3$ using the OSC:
if label = 1
$\mathbf{Z}_{cor} = \mathbf{X}_3 - \mathbf{X}_3 \cdot \mathbf{h}$
else $\mathbf{Z}_{cor} = \mathbf{X}_3$;

3. Proposed method

3.1. Main idea

Suppose the dataset $\mathbf{S}$ to be $(\mathbf{X}, \mathbf{Y})$, where $\boldsymbol{X} = (\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_N)^T \in \Re^{N \times d}$, $\boldsymbol{x}_i = (x_{i1}, \cdots, x_{id})$ and $\boldsymbol{Y} \in \Re^N$. N denotes the number of training samples and d is the feature dimension. The prediction function $f(\cdot)$ can be defined as:

$$\min_f \sum_{i=1}^{N} ||f(\mathbf{x}_i) - Y_i||_2^2 \tag{1}$$

The absolute prediction error can be calculated by

$$error_i = |f(\mathbf{x}_i) - Y_i|, i = 1, \cdots, N \tag{2}$$

We can imagine that a larger prediction error will be obtained by using the learned prediction function $f(\cdot)$ on some samples with different pattern from the pattern $\mathbf{S}$. Therefore, an unknown sample can be discriminated whether it belongs to the patterns from $\mathbf{S}$ or not by finding an appropriate threshold T. For example, if a sample (U, V) with different pattern from $\mathbf{S}$ is used as the input of $f(\cdot)$, we can obtain that $|f(\mathbf{U}) - V| > T$. This idea implies that the invariant information of the fixed dataset $\mathbf{S}$ has been memorized in the learned prediction function $f(\cdot)$. The new patterns that have different space distribution from $\mathbf{S}$ will have a conflict, such that it is easy to recognize all types of different patterns that are uncorrelated with the invariant information in $\mathbf{S}$.

In this work, this idea is motivated by the fact that there are many kinds of interferences in e-nose applications. From the viewpoint of experiments, it is impossible to collect all interferences samples. However, the target gases, as fixed patterns, can be recognized as invariant information, such that the proposed idea is reasonable for interference elimination. Motivated by the information invariance and the proposed idea, a pattern mismatch based interference elimination (PMIE) method is proposed. Note that it is "pattern mismatch" instead of "pattern match" which aims at recognizing the interference by the mismatch degree (i.e. prediction error), instead of recognizing the target gases.

3.2. Pattern mismatch based interference elimination (PMIE)

The proposed PMIE method includes two parts: interference discrimination and interference elimination. The details for each part have been presented as follows.

3.2.1. PMIE based interference discrimination

In the e-nose, the responses of temperature (x_{tem}), humidity (x_{hum}) and three MOS sensors (x_{s1}, x_{s2}, x_{s3}) are used to predict the response of the fourth sensor (x_{s4}). In the proposed PMIE method, the pattern vector $\mathbf{x}$ is described as $(x_{tem}, x_{hum}, x_{s1}, x_{s2}, x_{s3})$ and the predicted variable Y is described as x_{s4}. In this paper, the existing regression techniques such as regularized least square (RLS), back propagation neural network (BPNN) and support vector regression (SVR) have been used to learn the prediction function $f(\cdot)$, by assuming that $x_{s4} = f(x_{tem}, x_{hum}, x_{s1}, x_{s2}, x_{s3})$. We call them PMIE-RLS, PMIE-BP and PMIE-SVR, respectively. Note that in this paper, four gas sensors have been used as predicted variable respectively.

For the discrimination of an unknown sample $\mathbf{x}$, the *label* of $\mathbf{x}$ can be obtained by calculating the *error* between its predicted output $y = f(x_{tem}, x_{hum}, x_{s1}, x_{s2}, x_{s3})$ and the true value x_{s4}, as follows

$$label = \begin{cases} 1, & if \ error > T \\ 0, & if \ error \leq T \end{cases} \tag{3}$$

where the threshold T can be determined by ROC curve and misjudgment rate (MR). Note that the *label* "1" denotes the interference pattern (i.e. non-target gas) and "0" denotes target gas pattern.

Specifically, the interference discrimination process of the PMIE method is shown in Fig. 3. Note that the training process of the sensor prediction function $f(\cdot)$ in Fig. 3 is conducted by Dataset 1.

3.2.1.1. Threshold determination.
Dataset 2 was used to determine the threshold T for pattern mismatch. Specifically, the ROC (receiver operating characteristic) curve was used to test the performance of the prediction model, which was a popular index in reflecting the sensitivity and specificity of recognition. The area under the curve can reflect the overall recognition performance [17–19]. For a binary classification problem, four items such as true positive, true negative, false positive and false negative are included. In ROC, the true positive rate (TPR) is calculated by

$$TPR = \frac{TP}{TP + FN} \tag{4}$$

The false positive rate (FPR) is calculated by

$$FPR = \frac{FP}{FP + TN} \tag{5}$$

In Eq. (4) and (5), TP is the number of true positive samples, FN is the number of false negative samples, FP is the number of false positive samples, and TN is the number of true negative samples.

First, the prediction error vector of dataset 2 can be obtained by using the learned prediction function $f(\cdot)$, then the maximum and minimum errors denoted as E_{max} and E_{min} are obtained. The interval $[E_{min}, E_{max}]$ is divided into 100, such that there are 100 potential thresholds. For each threshold, the sensitivity (TPR, true positive rate) and specificity (FPR, false positive rate) can be calculated for ROC curve. Then, the misjudgment rate (MR) could be calculated for each threshold, in which the threshold with the smallest misjudgment rate is chosen as the final threshold T in Eq. (3).

3.2.2. PMIE based interference elimination
After interference discrimination, the proposed interference elimination algorithm is formulated as follows.

$$z_{cor} = \begin{cases} \mathbf{z} - \boldsymbol{h} \cdot \mathbf{z}, & label = 1 \\ \mathbf{z}, & label = 0 \end{cases} \tag{6}$$

where $\mathbf{z}$ is the sensor array response and $\mathbf{h}$ is the correction vector being determined. Here the orthogonal signal correction algorithm (OSC) is used to establish the signal correction matrix $\mathbf{h}$.
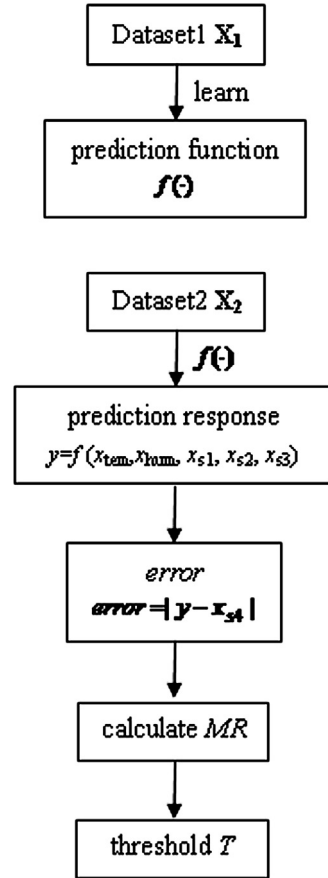
The OSC has been used to remove the irrelevant information (orthogonal part) in the projection vector from the spectral matrix [20–22]. In this paper, the interference information needed to

Table 4
Average training and testing prediction errors for the three models.

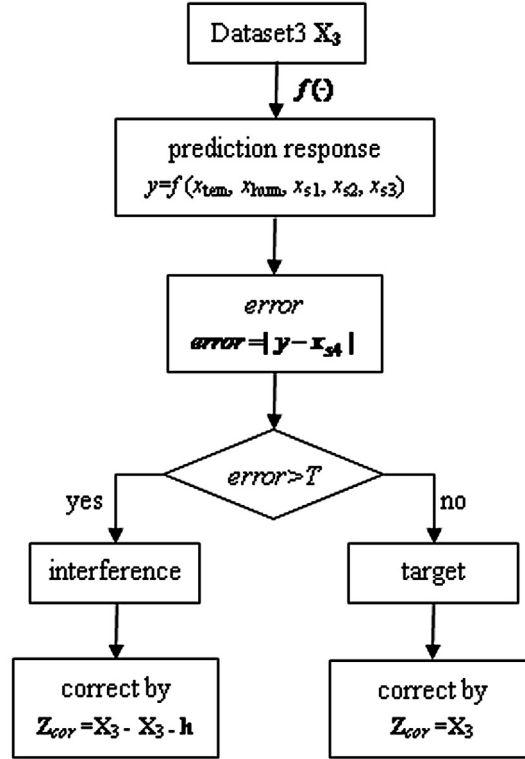| Predicted sensor	TGS2620 train/test error	TGS2602 train/test error	TGS2201A train/test error	TGS2201B train/test error
PMIE-RLS	0.0269/0.0242	0.0545/0.055	0.0475/0.0499	0.0213/0.0182
PMIE-BP	0.0106/0.0112	0.015/0.016	0.019/0.021	0.0171/0.02
PMIE-SVR	2.2e-5/2.0e-5	1.6e-5/1.4e-5	2.3e-5/1.7e-5	1.0e-5/9.4e-6

Fig. 4. Flowchart of PMIE algorithm.

remove is irrelevant to the target signal, thus the OSC algorithm can be used to remove it. Correspondingly, during the removal of interference information, the mean vector **R** of the nearest n real-time target samples is used as the projection vector and the real-time sensor interference signal **z** is used as the spectral matrix. The mean vector **R** of the nearest n real-time target samples **P** can be defined as:

$$R_i = \frac{1}{n}\sum_{j=1}^{n} P_{i,j}, \quad i = 1, 2, \cdots, d \quad (7)$$

where $\mathbf{R} = [R_1, \cdots, R_d]^T$, $\mathbf{P} = [\mathbf{P}_1, \cdots, \mathbf{P}_d]^T$, n is the length of the vector $\mathbf{P}_i$, and here, n is set to be 20. d is the feature dimension (i.e. the number of sensors). The correction vector h can be determined by the OSC algorithm.

3.2.3. PMIE algorithm

The training and testing stage of the PMIE algorithm have been illustrated in Tables 2 and 3, respectively. And the flowchart of the PMIE algorithm is shown in Fig. 4.

4. Results and discussion

4.1. Data preprocessing

Smooth filtering [11] and the vector standardization are used for preprocessing and normalization, respectively, in this paper. The filtered signal vector **X** can be calculated by

$$X(i) = \frac{\sum_{l=i}^{i+20-1} q_l - \max(q_i, \cdots, q_{i+20-1}) - \min(q_i, \cdots, q_{i+20-1})}{18}$$

$$i = 1, \cdots, n - 20 + 1 \quad (8)$$

where $\mathbf{q} = [q_1, q_2, \cdots, q_n]^T$ is a response sequence for one sensor, and the length is n. $\mathbf{X} = [X(i), \cdots, X(n - 20 + 1)]^T$ is the filtered signal sequence for one sensor. The length of the smoothing filter is set as 20. By subtracting the maximum and minimum values, part of the outliers may be removed, while the noise can be filtered out by averaging the remainder samples (which functions essentially as a low-pass filter).
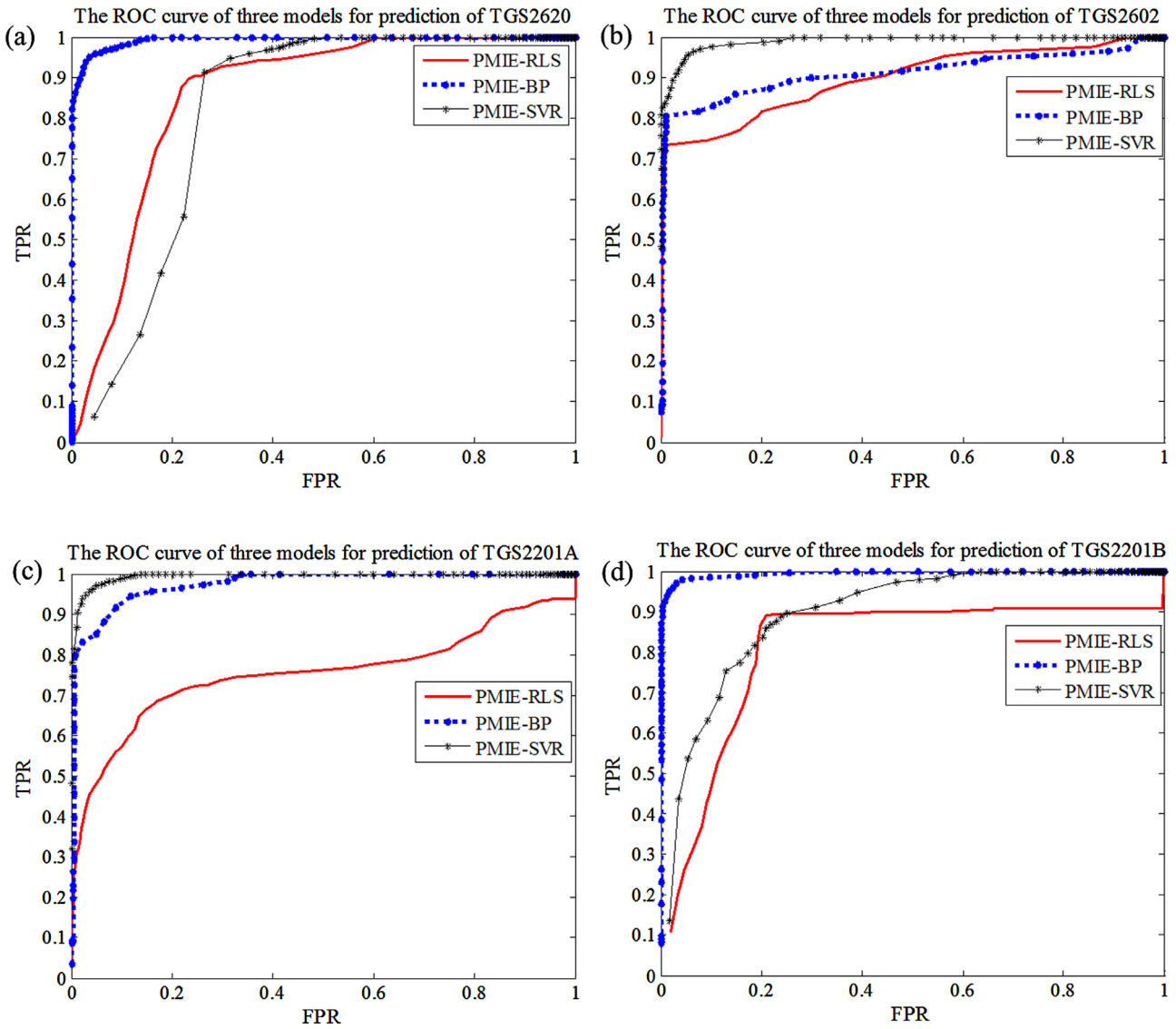
Fig. 5. ROC curves of the sensor TGS6220 (a), TGS2602 (b), TGS2201A (c) and TGS2201B (d) based on the three prediction models.

The vector standardization is used as a normalization method.

$$y_{ij} = \frac{x_{ij}}{\sqrt{\sum_i (x_{ij})^2}}, \quad i = 1, \cdots, 4; j = 1, \cdots, n \qquad (9)$$

It is worthy of noting that the vector standardization, as one kind of normalization method, makes the length of response vector of sensor array or the norm of vector to be per unit length, that is, i.e., $\|\mathbf{y}_j\| = 1$, with $\mathbf{y}_j = [y_{1j}, y_{2j}, y_{3j}, y_{4j}]^T$. This is a common used normalization method which diminishes the influence of concentration to some extent [23].

4.2. PMIE training on Dataset 1

Three models such as PMIE-RLS, PMIE-BP and PMIE-SVR are used in this paper. The average prediction errors of each sensor for the three models are listed in Table 4, and it is can be seen that the prediction accuracy of the PMIE-SVR model shows the best performance. However, the proposed PMIE method has less strict requirement on the prediction accuracy, as long as the boundary

points which discriminate between target gas response and interference gas response can be found.

4.3. Threshold analysis based on Dataset 2

Dataset 2 was used to determine the threshold. First, the prediction errors were obtained from dataset 2 for different prediction models, then the maximum error E_{max} and minimum error E_{min} can be achieved. The interval $[E_{min}, E_{max}]$ was divided into 100 thresholds uniformly. For each threshold, the sensitivity (TPR) and specificity (FPR) can be calculated. The ROC curves are shown in Fig. 5, which show that the areas under curve (AUC) of the ROC for three models are more than 0.5. The PMIE-BP model is better than PMIE-SVR and PMIE-RLS models for pattern prediction of TGS2620 and TGS2201B. The PMIE-SVR model is the best model for pattern prediction of TGS2602 and TGS2201A. In general, the PMIE-BP model and PMIE-SVR model have their own advantages in discrimination, and the PMIE-RLS is the worst due to the linear regression nature.

Second, the misjudgment rate can be calculated for each threshold. The threshold with the smallest misjudgment rate is used as the threshold for interference discrimination. Relationships between
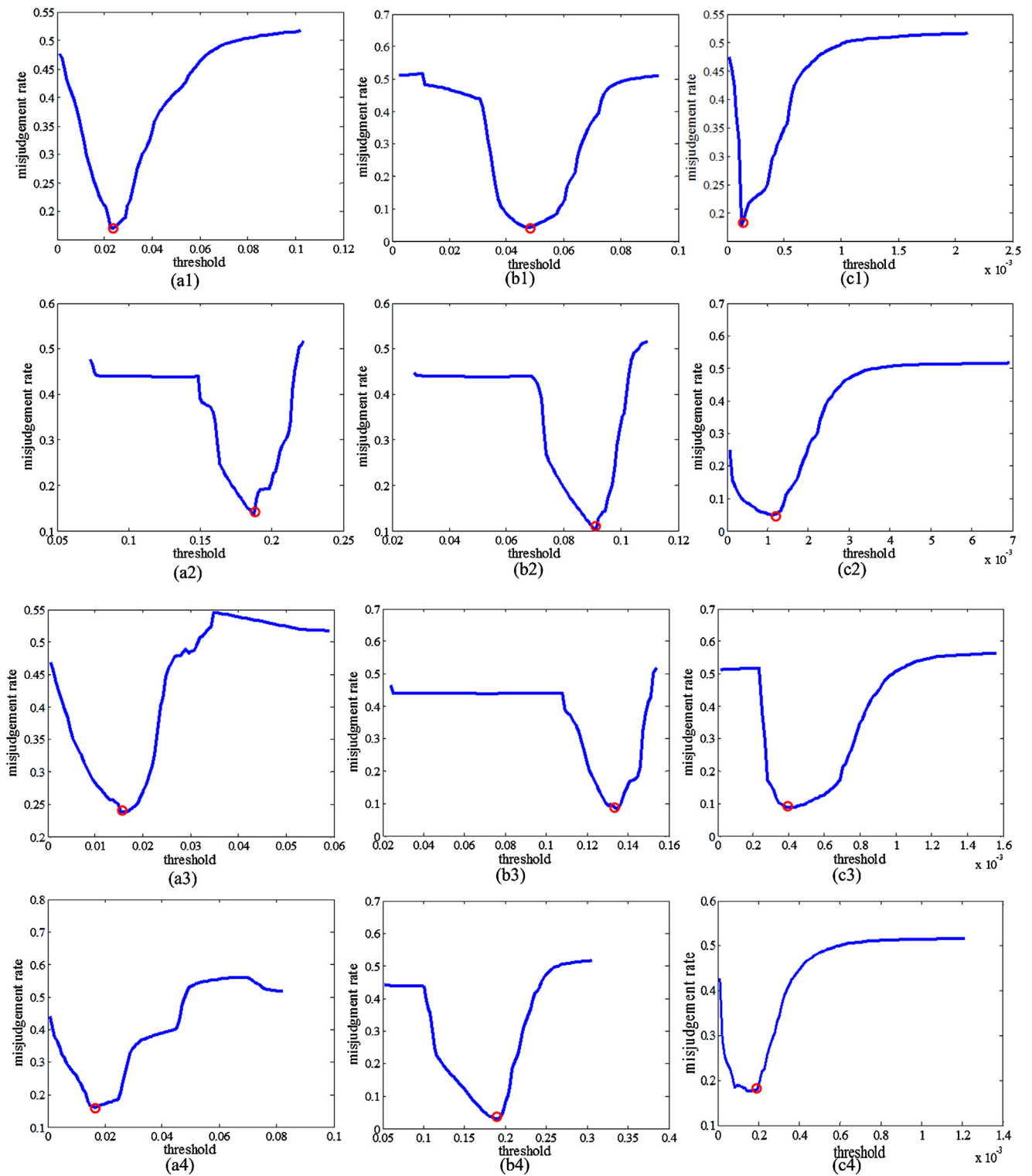
Fig. 6. Misjudgment rate (MR) with threshold (T) for the four sensors by using PMIE-RLS (a1–a4), PMIE-BP (b1–b4), and PMIE-SVR (c1–c4), respectively.

Table 5
Threshold (T) and misjudgment rate (MR) of four sensors for the three models.

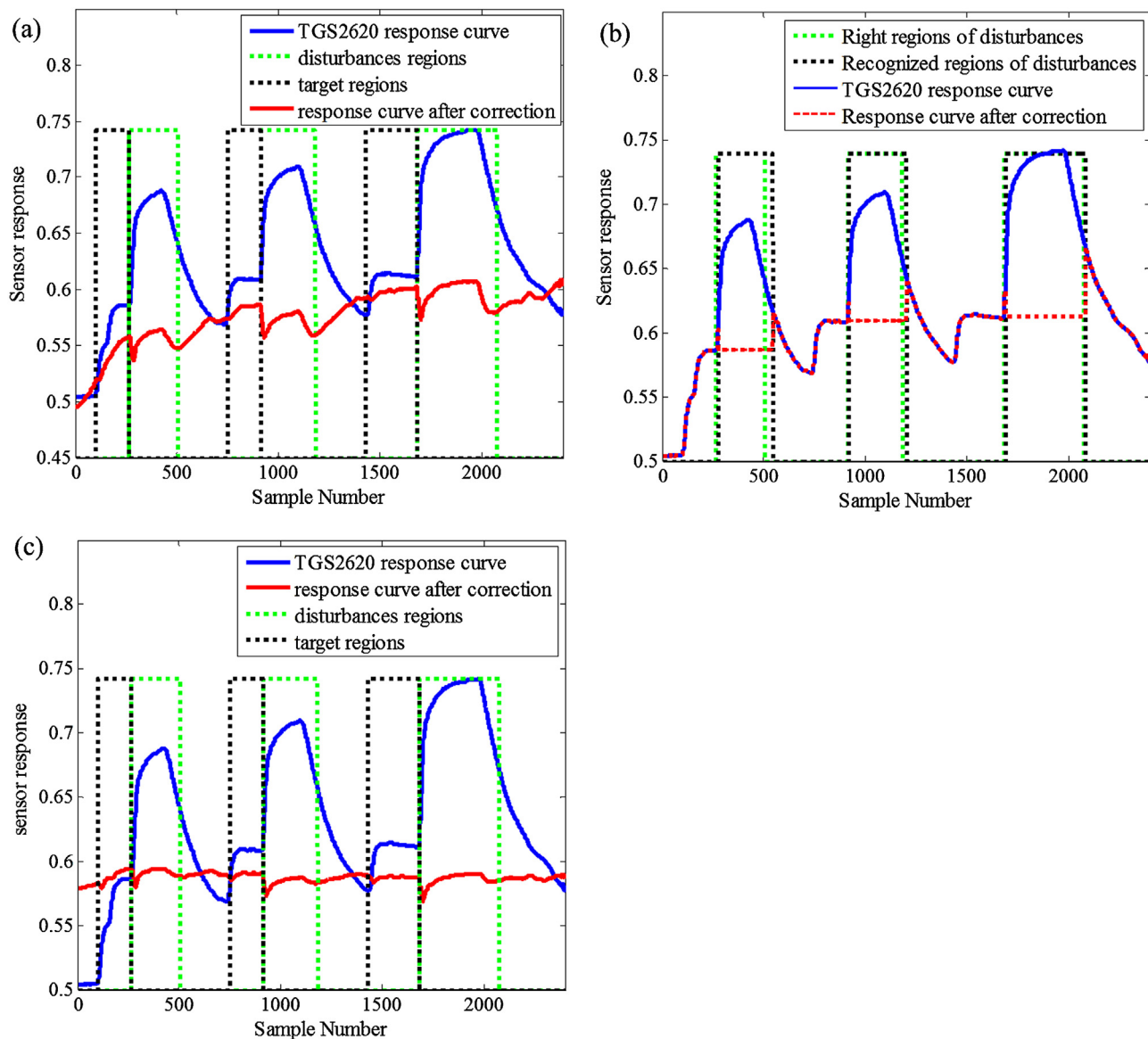| model	PMIE-RLS		PMIE-BP		PMIE-SVR	
	T	MR	T	MR	T	MR
TGS2620	0.024	17.38%	0.0464	7.46%	0.00015	18.75%
TGS2602	0.188	13.58%	0.0905	10.54%	0.0012	5.08%
TGS2201A	0.0183	24.80%	0.1342	8.58%	0.0004	9.67%
TGS2201B	0.0189	16.90%	0.1802	4%	0.00018	17.83%

Fig. 7. Curves of sensor TGS2620 after correction by ICA algorithm (a), OSC with PMIE-BP model discrimination (b) and OSC algorithm (c).

threshold and misjudgment rate are shown in Fig. 6, from which the threshold and misjudgment rate of each model for each sensor can be observed. The red point represents the threshold point. The details for each sensor are described in Table 5.

In terms of the misjudgment rate, the PMIE-BP model with respect to the TGS2201B sensor as the dependent variable in PMIE has the smallest misjudgment rate for interference discrimination.

In summary, the training process of the discrimination model $f(\cdot)$ based on Dataset 1 and the search for the best threshold T based on Dataset 2 are described in this section. The obtained $f(\cdot)$ and T will be tested in the following section by using Dataset 3.

4.4. Interference discrimination result on Dataset 3

4.4.1. Discrimination result based on single model

In this section, the prediction results of three models for each sensor are described in Table 6. It can be seen that for each sensor, the misjudgment rates of PMIE-BP model and PMIE-SVR model are smaller. Therefore, the PMIE-BP model and PMIE-SVR model are more suitable for interference discrimination.

Table 6
Misjudgment rate (*MR*) of four sensors for the three models.

model	PMIE-RLS MR	PMIE-BP MR	PMIE-SVR MR
TGS2620	17.75%	3.63%	12.04%
TGS2602	6%	3.91%	2.33%
TGS2201A	28.40%	3.83%	6.70%
TGS2201B	34.70%	8.38%	8.50%

4.4.2. Discrimination result based on multiple models

For each model, the responses of four sensors are predicted, respectively, i.e. $S_1 = f_1(S_2, S_3, S_4)$, $S_2 = f_2(S_1, S_3, S_4)$, $S_3 = f_3(S_1, S_2, S_4)$ and $S_4 = f_4(S_1, S_2, S_3)$. So there are four prediction functions for each model (e.g. SVR). In order to prevent the recognition error caused by a single sensor prediction, it uses multiple sensor predictions to discriminate between target signals and interference signals in the same model. As shown in Table 7, k ($k = 1, 2, 3$) denotes that a signal is discriminated as an interference response when the signal is predicted as interference by k prediction functions. Table 7 illustrates that the misjudgment rate is smallest with $k = 3$ under the same model.

Table 7
Comparison of the misjudgment rate (*MR*) between the proposed PMIE and method [11].

Method k	The proposed method									Method [11]		
	PMIE-RLS			PMIE-BP			PMIE-SVR			RLS	BP	SVR
	1	2	3	1	2	3	1	2	3			
MR	42.10	15.25	10.45	5.67	3.41	3.25	5.41	5.62	5.29	6.55	5.75	5.08

The PMIE method proposed in this paper is different with the pure classification based method [11]. It only relies on the target samples when it is set up, and it is irrelevant to interference gases. So the method is effective to all interference gases. However, the method in Ref. [11] needs some interference samples to train classifier, so it depends on the type of interference gas to some extent. Thus it is effective to limited types of interference. And Table 7 shows that the PMIE-BP method is superior to the method in [11] in discrimination performance.

4.5. PMIE based interference elimination result

The OSC algorithm was used to correct the interfered responses. The curves of TGS2260 responses which were discriminated by PMIE-BP prediction model and corrected by OSC are listed in Fig. 7(b), wherein we can find that the misjudgment positions are the boundaries between the target gases and interferences. But the boundary problem is still an unsolved issue, which is a cost-sensitive problem.

The independent component analysis algorithm (ICA) and orthogonal signal correction algorithm (OSC) can be used in components correction [24–27]. In the e-nose based on MOS sensor array, they are used to solve the interference problem and drift problem [28–30] caused by sensor aging. For comparison, the ICA have been utilized for interference elimination. The TGS2620 response curve with interference elimination is shown in Fig. 7(a). Additionally, the single OSC method also have been utilized, which can remove the information that is irrelevant to target response. The corrected curve of TGS2620 is shown in Fig. 7(c). From the results, it is can be concluded that the proposed PMIE method can successfully discriminate and eliminate the interferences in real-time application. This also implies that ICA or OSC may not effectively eliminate the interferences that are independent of target gases in real-time application from their principle. Therefore, for real-time use of an e-nose in real-world scenarios, the proposed method can well address the issue in e-nose community.

4.6. Assumptions and limitations of PMIE model

Assumptions and limitations of the model are discussed as follows:

4.6.1. Assumption
For the PMIE model, we assume that the target gas concentration remains constant when the interference gas exists.

4.6.2. Limitations
1) The threshold *T* used to discriminate the signal attribute should be very accurate in the model. If the target signal and the interference signal are very similar, the probability of wrong discrimination will be larger.
2) During the process of interference elimination, the interference information orthogonal to the target signal is removed, but it cannot guarantee that all the interference information is removed. Therefore the interference is not completely eliminated while it can be suppressed to a large extent.

3) For the model in this paper, the signal is discriminated as one of the two kinds: either target signal or interference. However, in the future, we will consider multi-label prediction, that is, multiple kinds of gases can be discriminated simultaneously.

5. Conclusion

In this paper, a pattern mismatch based interference elimination (PMIE) method is proposed for addressing the key and unsolved issue of interferences in e-nose. This method consists of two parts: interference discrimination and interference elimination. The discrimination model of PMIE only relies on target gas samples, and it is built independently of specific interference gases, such that the PMIE method is effective to all interference gases. By comparing with ICA and OSC in experiments, the proposed PMIE method shows the best results. It means that the interference problems caused by non-target gases in e-nose was satisfactorily solved by the proposed PMIE method.

In the future, the boundary problem of discrimination based on a well-trained threshold should be further studied. Additionally, more effective mismatch based discrimination model may also be an interesting research direction.

References

[1] S.M. Scott, D. James, Z. Ali, Data analysis for electronic nose systems, Microchim. Acta. 156 (2007) 183–207.
[2] L. Zhang, F. Tian, Performance study of multilayer perceptrons in a low-Cost electronic nose, IEEE Trans. Instrum. Meas. 63 (2014) 1670–1679.
[3] X. Tian, Y. Yin, H. Liu, Research on artificial olfactory sensor technology for liquor identification, Food Sci. 2 (2004) 29–32.
[4] B. Mumyakmaz, A. Özmen, M.A. Ebeoğlu, C. Taşaltın, İ. Gürol, A study on the development of a compensation method for humidity effect in QCM sensor responses, Sens. Actuators B: Chem. 1 (2010) 277–282.
[5] K.R. Kashwan, M. Bhuyan, Robust electronic-nose system with temperature and humidity drift compensation for tea and spice flavour discrimination, in: Proceedings of the Sensors and the International Conference on New Techniques in Pharmaceutical and Biomedical Research, Kuala Lumpur, Malaysia, 5–7 September, 2005, pp. 154–158, http://dx.doi.org/10.1109/ASENSE.2005.1564528.
[6] J.W. Gardner, E.L. Hines, F. Molinier, P.N. Bartlett, T.T. Mottram, Prediction of health of dairy cattle from breath samples using neural network with parametric model of dynamic response of array of semiconducting gas sensors, Sci. Meas. Technol. IEE Proc. 2 (1999) 102–106.
[7] X.L. Xu, J.N. Qiu, C. Chen, A study on local sensor fusion of wireless sensor networks based on the neural network, Mach. Learn. Cybern. 7 (2008) 4045–4050.
[8] J.F. Shi, H.B. Tang, H.Y. Gong, Application of wavelet neural network and multi-sensor data fusion technique in intelligent sensor, World Congr. Intell. Control Autom. (2008) 1114–1117.
[9] T.A. Emadi, C. Shafai, M.S. Freund, D.J. Thomson, D.S. Jayas, N.D.G. White, Development of a polymer-based gas sensor—humidity and CO_2 sensitivity, Microsyst. Nanoelectron. Res. Conf. (2009) 112–115.

[10] C. Di Natale, E. Martinelli, A. D'Amico, Counteraction of environmental disturbances of electronic nose data by independent component analysis, Sens. Actuators B: Chem. 82 (2002) 158–165.

[11] L. Zhang, F. 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: Phys. 201 (2013) 254–263.

[12] J. Feng, F. Tian, J. Yan, Q. He, Y. Shen, L. Pan, A background elimination method based on wavelet transform in wound infection detection by electronic nose, Sens. Actuators B: Chem. 157 (2011) 395–400.

[13] F. Tian, J. Yan, S. Xu, J. Feng, Q. He, Y. Shen, P. Jia, Background interference elimination in wound infection detection by electronic nose based on reference vector-based independent component analysis, Inform. Technol. J. 7 (2012) 850–858.

[14] N.G. Yee, G.G. Coghill, Factor selection strategies for orthogonal signal correction applied to calibration of near-infrared spectra, Chemom. Intell. Lab. Syst. 67 (2003) 145–156.

[15] J. Feng, F. Tian, P. Jia, Q. He, Y. Shen, S. Fan, Improving the performance of electronic nose for wound infection detection using orthogonal signal correction and particle swarm optimization, Sens. Rev. 34 (2014) 389–395.

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

[17] X. Zhang, X. Li, Y. Feng, Z. Liu, The use of ROC and AUC in the validation of objective image fusion evaluation metrics, Signal Process. 115 (2015) 38–48.

[18] V. Nykänen, I. Lahti, T. Niiranen, K. Korhonen, Receiver operating characteristics (ROC) as validation tool for prospectivity models—a magmatic Ni–Cu case study from the Central Lapland Greenstone Belt, Northern Finland, Ore Geol. Rev. 71 (2015) 853–860.

[19] M. Thomas, K. De Brabanter, J.A.K. Suykens, B. De Moor, Predicting breast cancer using an expression values weighted clinical classifier, BMC Bioinf. 15 (2014) 1–11, http://dx.doi.org/10.1186/s12859-014-0411-1.

[20] S. Wold, H. Antti, F. Lindgren, J. Öhman, Orthogonal signal correction of near-infrared spectra, Chemom. Intell. Lab. Syst. 44 (1998) 175–185.

[21] Z. Talebpour, R. Tavallaie, S.H. Ahmadi, A. Abdollahpour, Simultaneous determination of penicillin G salts by infrared spectroscopy: evaluation of combining orthogonal signal correction with radial basis function-partial least squares regression, Spectrochim. Acta Part A: Mol. Biomol. Spectrosc. 76 (2010) 452–457.

[22] L. Laghi, A. Versari, G.P. Parpinello, D.Y. Nakaji, R.B. Boulton, FTIR spectroscopy and direct orthogonal signal correction preprocessing applied to selected phenolic compounds in red wines, Food Anal. Methods 4 (2011) 619–625.

[23] J.W. Gardner, E.L. Hines, H.C. Tang, Detection of vapours and odours from a multisensor array using pattern-recognition techniques Part 2. Artificial neural networks, Sens. Actuators B: Chem. 9 (1992) 9–15.

[24] D.J. Bouveresse, A. Moya-González, F. Ammari, D.N. Rutledge, Two novel methods for the determination of the number of components in independent components analysis models, Chemom. Intell. Lab. Syst. 112 (2012) 24–32.

[25] S. Balasubramanian, S. Panigrahi, C.M. Logue, C. Doetkott, M. Marchello, J.S. Sherwood, Independent component analysis-processed electronic nose data for predicting Salmonella typhimurium populations in contaminated beef, Food Control 19 (2008) 236–246.

[26] T. Aguilera, J. Lozano, J.A. Paredes, F.J. Alvarez, J.I. Suarez, Electronic nose based on independent component analysis combined with partial least squares and artificial neural networks for wine prediction, Sensors (Basel) 6 (2012) 8055–8072.

[27] M. Padilla, A. Perera, I. Montoliu, A. Chaudry, K. Persaud, S. Marco, Drift compensation of gas sensor array data by orthogonal signal correction, Chemom. Intell. Lab. Syst. 100 (2010) 28–35.

[28] L. Zhang, F. 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: Chem. 182 (2013) 71–79.

[29] M. Holmberg, F.A.M. Davide, C. Di Natale, A. D'Amico, F. Winquist, I. Lundström, Drift counteraction in odour recognition applications: lifelong calibration method, Sens. Actuators B: Chem. 42 (1997) 185–194.

[30] L. Zhang, D. Zhang, Domain adaptation extreme learning machines for drift compensation in E-nose systems, IEEE Trans. Instrum. Meas. 64 (2015) 1790–1801.

Biographies

Fengchun Tian received his 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.

Zhifang Liang received her Bachelor degree in Computer Science and Technology in 2013 from Shanxi Normal University, China. Since September 2014, she has been studying for a Ph.D. degree in intelligent signal processing. Her research interests include electronic nose and intelligent algorithm.

Lei Zhang received his Ph.D. degree in Circuits and Systems from the College of Communication Engineering, Chongqing University, Chongqing, China, in 2013. He was selected as a Hong Kong Scholar in China in 2013 and worked as a Post-Doctoral Fellow with The Hong Kong Polytechnic University, Hong Kong. Now, he is a Distinguished Research Fellow in Chongqing University. He has authored more than 40 scientific papers in electronic nose, machine olfaction, pattern recognition, and machine learning. His current research interests include machine learning, pattern recognition, artificial olfactory system, and nonlinear signal processing in Electronic Nose. Dr. Zhang was a recipient of Outstanding Doctoral Dissertation Award of Chongqing, China, in 2015, Hong Kong Scholar Award in 2014, Academy Award for Youth Innovation of Chongqing University in 2013 and the New Academic Researcher Award for Doctoral Candidates from the Ministry of Education, China, in 2012.

YanLiu received her Bachelor degree in Information Engineering in 2014 from Chengdu Polytechnic University, China. Since September 2014, she has been studying for a MS degree in modern signal processing. Her research interests include electronic nose and circuit design.

Zhenzhen Zhao received her MS degree in Electronic and Communication from Chongqing University, Chongqing, China in 2012. She is currently pursuing the Ph.D. degree in the same university. Her research interests include optical gas sensing techniques and optical signal processing. She is currently working on the optical gas sensing system for electronic nose using wide spectrum light source.