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

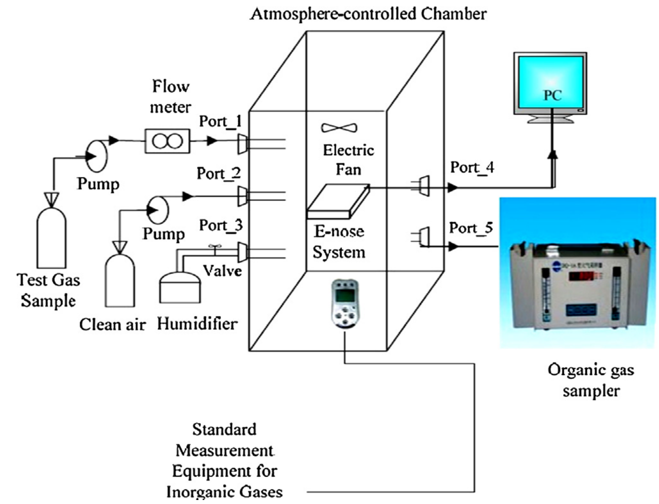

Fig. 1. Experimental platform of E-nose.

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.

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),

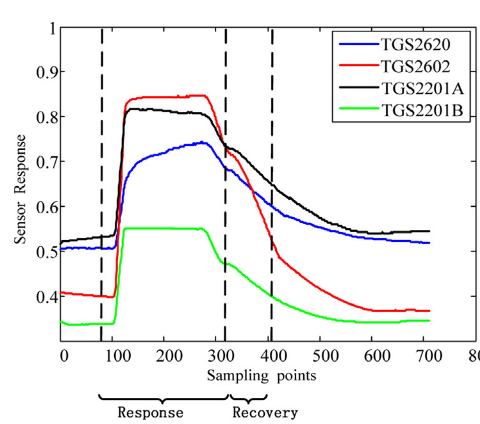

Fig. 2. Response of sensors array to formaldehyde.