Available online at www.sciencedirect.com
ScienceDirect
Sensors and Actuators B 120 (2007) 584–594
SENSORS AND ACTUATORS B CHEMICAL
www.elsevier.com/locate/snb

Simultaneous estimation of odor classes and concentrations using an electronic nose with function approximation model ensembles

Gao Daqi [a,b,*], Chen Wei [a]

[a] Department of Computer Science, East China University of Science and Technology, Shanghai 200237, China
[b] State Key Laboratory of Bioreactor Engineering, East China University of Science and Technology, Shanghai 200237, China

Received 13 December 2005; received in revised form 14 March 2006; accepted 15 March 2006
Available online 24 April 2006

Abstract

This paper sets up a practical electronic nose for simultaneously estimating many kinds of odor classes and concentrations. Mathematically, such simultaneous estimation problems can be regarded as multi-input/multi-output (MIMO) function approximation problems. After decomposing an MIMO approximation task into multiple many-to-one tasks, we can use multiple many-to-one approximation model ensembles to implement them one after another. A single approximation model may be a multivariate logarithmic regression, a quadratic multivariate logarithmic regression, a multilayer perceptron, or a support vector machine. An ensemble is made up of the above four models, represents a special kind of odor, and realizes the relationship between sensor array responses and the represented odor concentrations. Naturally, all members in the ensemble are trained only by the samples from the represented odor. The real outputs of ensembles are the average predicted concentrations and the relative standard deviations (R.S.D.s). The ensemble with the smallest R.S.D. finally gives the label and concentration of an odor sample, which can be looked upon as the use of the average and the minimum combination rules. The predicted results for four kinds of fragrant materials, ethanol, ethyl acetate, ethyl caproate, and ethyl lactate, 21 concentrations in total, show that the proposed approximation model ensembles and combination strategies with the electronic nose are effective for simultaneously estimating many kinds of odor classes and concentrations.

Keywords: Simultaneous estimations; Odor classes; Odor concentrations; Function approximation models; Combination strategies; Electronic noses

1. Introduction

Compared with a man of rich experience, today's electronic noses are quite limited in capabilities [1,2]. Up to now, electronic noses can only recognize limited kinds of odors [1–5], or estimate limited kinds of concentrations of simple odors [6,7]. What is serious is that there are few reports in the literature for electronic noses to simultaneously determine many kinds of odor classes and concentrations [8–11], which is far from man's expectation. One of the main obstacles is that the performances of the existing classification and function approximation methods are not so ideal as expected [1,2].

To simultaneously estimate many kinds of odor classes and strengths is really to simultaneously decide which class an odor sample belongs to and what size its concentration is. Since electronic noses employ sensor arrays to sense odors, such a problem can be regarded as either a multivariate classification or a multivariate function approximation problem mathematically. Therefore, there are the following two solutions:

(A) Take a simultaneous estimation problem for a multivariate classification problem, and employ either a many-to-many or multiple many-to-one classifiers to solve it. An odor concentration point is regarded as a class. To use a many-to-many classifier really means to look upon a multi-class problem as a whole. If there are n_1 kinds of odors and n_2 concentrations in each, $n_1 \times n_2$ output units are needed. Some of shortcomings of multi-input/multi-output classifiers are of low classification accuracy and long learning time when there exist too many classes. On the other hand, to use multiple many-to-one classifiers is to decompose a multi-class problem into multiple two-class problems, and make a single-output classifier on behalf of a concentration-fixed odor. The deficiency to do like that is that too many classifiers are needed, which is especially impractical when

* Corresponding author. Tel.: +86 21 6424 6970; fax: 86 21 6425 2830.
E-mail address: gaodaqi@8163.net.cn (G. Daqi).

0925-4005/$ – see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.snb.2006.03.017

there is a need to estimate a great quantity of strengths in many kinds of odors.

(B) Regard a simultaneous estimation problem as a multivariate function approximation problem, and use either a many-to-many or multiple many-to-one approximation models to solve it. The two methods are not ideal in effect, because a single approximation model does not know to say "No". In other words, an output unit of approximation models always gives a predicted value for an input, no matter how ridiculous it is. An awful nuisance is that two output units may give self-contradictorily predicted results for an odor sample. What can we do if an output unit takes a sample for 100 ppm of methanol, but another takes it for 1000 ppm of ethanol? Obviously, a single kind of function approximation models is unable to cope with the simultaneous estimation problems of many kinds of odor classes and concentrations. For a single sample, it is meaningless to consider which of two output units is of higher predicted accuracy.

The concept of classifier combinations was presented in 1990s [12–15], because the capabilities of a single type of classifiers are limited after all. For solving the large-scale learning problems, task decompositions and modular classifiers are also proposed [16,17]. In theory, the classifier combination strategies can be introduced into approximation model ensembles. But in practice, hardly can any one of them be used directly. In order to simultaneously estimate many kinds of odor classes and concentrations, this paper proposes a new kind of function approximation model ensembles and combination strategies. Our main ideas are as follows. An approximation model may be a multivariate logarithmic regression (MVLR) [3], a quadratic multivariate logarithmic regression (QMVLR) [18], a multilayer perceptron (MLP) [18,19], or a support vector machine (SVM) [20–22]. An ensemble is made up of the four kinds of function approximation models, and represents a specified odor. All members in an ensemble are trained only by the samples from the represented odor. If a sample indeed belongs to a certain represented odor, the viewpoints of all members in the representative ensemble are quite identical, otherwise quite divergent. The ensemble with the most identical viewpoints will finally label an odor sample, and the average predicted value of all experts in the ensemble determines the sample's concentration. Here, "the most identical viewpoint" is scaled by the relative standard deviation (R.S.D.) of predicted concentrations of a sample given by all experts in an ensemble. It is meaningless to only consider the outputs or predicted accuracies of a single approximation model ensemble. The above ideas will be explained in theory and verified by the experiments later.

This paper is organized as follows. Section 2 goes into details on the structure, principle and operating process of a practical electronic nose. The task decomposition method, the structures of function approximation components as well as their combination strategies, are formulated in Section 3. In Section 4, the presented method is applied to the simultaneous estimation of four kinds of fragrant materials, namely ethanol, ethyl acetate, ethyl caproate and ethyl lactate, 21 kinds of concentrations altogether. Finally, Section 5 comes to our conclusions.

2. Experimental

Fig. 1 is the schematic diagram of an improved electronic nose [8,9], which is mainly made up of a test box, a thermostatic cup, and a personal computer (PC). The test box consists of a gas sensor array, a three-channel high-precision direct constant power source, a gas sampling and flow control unit, a waste gas exhauster, a thermostatic unit, etc. The sensor array is composed of 16 TGS sensors, namely TGS800, 812, 813, 816, 821, 822, 823, 824, 825, 826, 830, 831, 832, 842, 880, 883 T, all provided by Figaro Engineering Inc., Japan. The array is installed within a 200 ml circular chamber, 140 mm in average diameter. A 16-bit high-precision data acquisition card is inserted in the expansion slot of the PC motherboard.

Fig. 2 charts the changes of gas flows and senor responses in a sampling procedure, which includes rough and fine restorations,

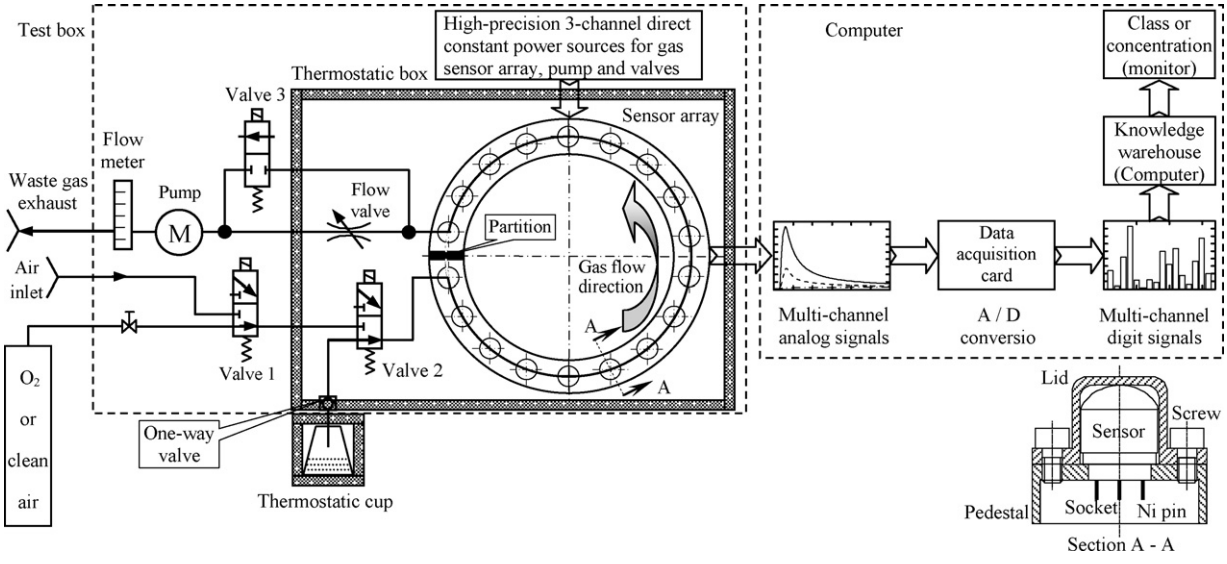

Fig. 1. Schematic diagram of a practical electronic nose.

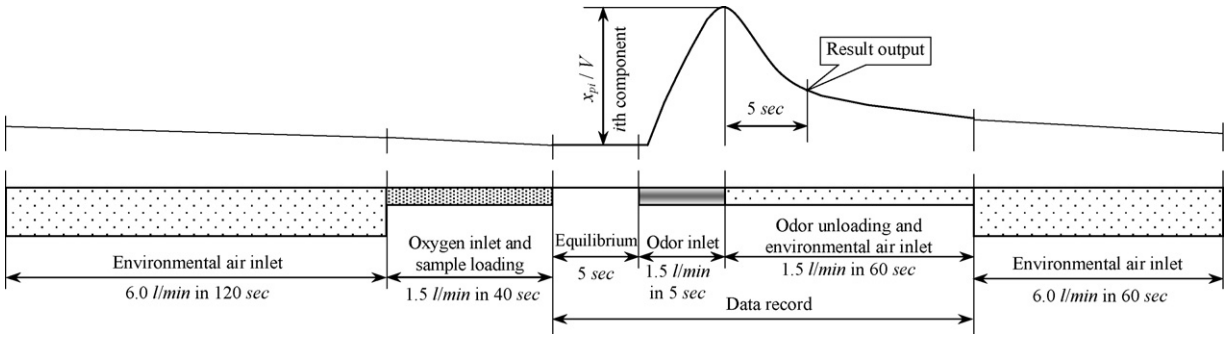

Fig. 2. Changes of sensor responses (upper) and gas flows (below) in a sampling procedure.

equilibrium, sampling, and rinse in small and large flows. The on–off orders of valves, rates of gas flows and durations are given in Table 1. In a 290 s work cycle, only the sensor responses during equilibrium, sampling, and small-flowed rinse are recorded, which lasts 70 s in total. The maximum response of a sensor, namely x_{pi} in Fig. 2, is used as a feature component.

The working principle of the electronic nose is as follows. A gas or headspace vapor sample waiting for measurement is drawn in the test box with a miniature diaphragm pump. The odor skims the sensitive membranes of sensors in the speed of 62.5 mm/s, and then is exhausted in the air through the outlet. During the flow of gas, the precise measuring unit records the responses of sensors. As a result, we get a 16-dimensional response vector in voltage form.

In order to ensure the same measurement condition, the following steps are adopted:

(A) Heating and temperature preservation of odor samples. A 30 ml liquid or solid sample is first poured in a 250 ml volumetric flask, heated by a semiconductor chip, and then kept in a thermostatic cup for 45 min. The cup is wrapped with heat insulation materials, and the thermostatic accuracy is $55 \pm 0.1\,^{\circ}C$.
(B) Heat preservation of key parts. The sensor array, the inlet pipes, the inlet electromagnetic valve, and the flow valve are all kept in a thermostatic box. The test box is wrapped with heat insulation materials and heated by a resistance heater. The thermostatic accuracy is $55 \pm 0.1\,^{\circ}C$.
(C) Calibration of sensor array. The array is purged by O_2 or dry air in the speed of 62.5 mm/s for 40 s before sampling.
(D) Control of sampling flow and volume. The sampling flow is controlled by an accurate flow valve in the range of 1.5 ± 0.01 l/min, and the sampling volume mainly by an accurate time relay in the range of 50 ± 1.0 ml. Once the volume of sample amounts to 50 ml, the electromagnetic valves will be switched on or off.

3. Function approximation models and combination strategies

3.1. Approximation model ensembles

Does a many-to-one function approximation model trained with the samples from odor j have an output for a sample from odor k? Yes! And the reverse is true. The embarrassing situation shows that a single type of function approximation model itself cannot simultaneously determine many kinds of odor classes and concentrations. So we turn to function approximation model ensembles as well as combination strategies for help [9]. n ensembles are representative of n kinds of odors, one for one. An ensemble is composed of several experts, and on behalf of a specified kind of odor. If sample $\boldsymbol{x}_p$ is indeed from odor j, the outputs of all members in ensemble j will simultaneously be close in on a predicted concentration; otherwise quite divergent. The most consistent panel ought to be taken to finally determine the class label and concentration of $\boldsymbol{x}_p$, as shown in Fig. 3.

A lot of experiments show that the relationships between sensor responses and odor strengths are close to logarithmic [1,2], thus MVLR models are suitable candidates to approximate them. Since the changes of logarithmic curves are not fiercely undulate [1–7], it is also suitable to use single-hidden-layer perceptrons with sigmoid activation functions (SAFs) and SVMs with radial basis function (RBF) kernels to execute such function approximation tasks. In addition, in consideration of measurement errors and environmental factors, the logarithmic curves are not wholly linear, so that QMVLR models are good approx-

Table 1
On–off order of electromagnetic valves and rate of gas flow

State	Valve 1	Valve 2	Valve 3	Rate of flow (l/min)	Duration (s)
Rough restoration	Off	On	On	6.0	120
Fine restoration	On	Off	Off	1.5	40
Equilibrium	Off	Off	Off	0.0	5
Sampling	Off	On	Off	1.5	5
Rinse with small flow	Off	On	Off	1.5	60
Rinse with large flow	Off	On	On	6.0	60

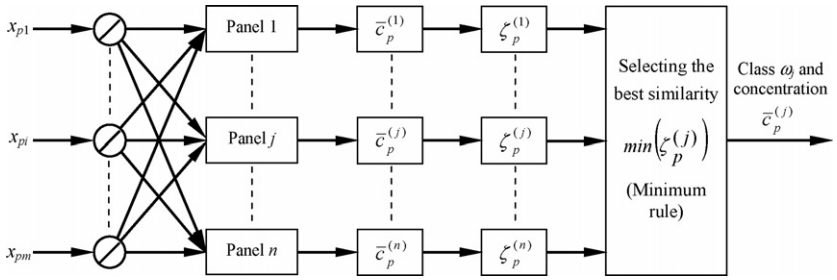

Fig. 3. Block diagram of expert ensembles.

imation tools. Therefore, this paper selects MVLRs, QMVLRs, MLPs with SAFs, and SVMs with RBF kernels to play the roles of experts, as shown in Fig. 4. Now there are four experts in each ensemble.

Let $\boldsymbol{x}_p = (x_{p1}, x_{p2}, \ldots, x_{pm})^{\mathrm{T}} \in R^m$ be the preprocessed response vector of a sensor array to the pth odor sample. The detailed approximation model ensembles and combination strategies are as follows:

(1) Decompose a multi-input multi-output approximation task f: $\boldsymbol{X} \to \boldsymbol{D}$ into multiple multi-input single-output tasks $f : \boldsymbol{X}^{(j)} \to \boldsymbol{d}_k^{(j)}$. Here $\boldsymbol{X} \in R^{N\times m}$, $\boldsymbol{X}^{(j)} \in R^{N_j\times m}$, $\boldsymbol{D} \in R^{N\times n}$, $\boldsymbol{d}_k^{(j)} \in R^{N_j\times 1}$, N is the total number of learning samples, and N_j only the number of training samples from odor j. In fact, N_j is far smaller than N.

(2) Use an MVLR, an QMVLR, an MLP, and an SVM, all with the m-to-1 structures, to form an ensemble j to fit the m-to-1 curve $f : \boldsymbol{X}^{(j)} \to \boldsymbol{d}_k^{(j)}$. In the training stage, all the four members in panel j learn the same input matrix $\boldsymbol{X}^{(j)}$, but their target outputs $\boldsymbol{d}_k^{(j)}$, namely the preprocessed concentration values, are not certainly the same.

(3) Predetermine the concentration of $\boldsymbol{x}_p$ by four experts in ensemble j according to the average combination rule:

$$\bar{c}_p^{(j)} = \frac{1}{4}\sum_{k=1}^{4} c_{pk}^{(j)} = \frac{1}{4}\sum_{k=1}^{4} \phi^{-1}(y_{pk}^{(j)}) \tag{1}$$

Here, $y_{pk}^{(j)}$ is the real output of expert k, $c_{pk}^{(j)}$ the predicted concentration of $\boldsymbol{x}_p$ given by expert k in panel j, and ϕ^{-1} expresses an inverse operation. The similar degree of $\boldsymbol{x}_p$ to odor j given by panel j is

$$\zeta_p^{(j)} = \frac{1}{\bar{c}_p^{(j)}}\sqrt{\frac{1}{4-1}\sum_{k=1}^{4}\left(c_{pk}^{(j)} - \bar{c}_p^{(j)}\right)^2} \times 100\,\% \tag{2}$$

In fact, $\zeta_p^{(j)}$ is similar to the R.S.D. of the average predicted concentration of $\boldsymbol{x}_p$ given by all experts in ensemble j.

(4) Determine the final label and concentration of $\boldsymbol{x}_p$ with the help of the minimum combination rule:

$$\boldsymbol{x}_p \in \omega_j \quad \text{and} \quad c_p = \bar{c}_p^{(j)} \quad \text{if } \zeta_p^{(j)} = \min_{1\le k\le n}(\zeta_p^{(k)})$$
$$j = 1, 2, \ldots, n \tag{3}$$

The real outputs of panel j for $\boldsymbol{x}_p$ are the similar degree $\zeta_p^{(j)}$ to the represented odor j as well as the average predicted concentration $\bar{c}_p^{(j)}$, as shown in Fig. 4.

3.2. Multivariate logarithmic regression models

Through learning, MVLR j is really to realize the many-to-one mapping $f : \boldsymbol{X}^{(j)} \to \boldsymbol{d}_1^{(j)}$ in order to minimize the root-mean-square (RMS) error $\varepsilon_1^{(j)} = \sqrt{1/2N_j}||\boldsymbol{d}_1^{(j)} - \boldsymbol{y}_1^{(j)}||_2$. Suppose the concentration of distilled water is 1.0×10^{-7}, namely 0.1 ppm, the real output vector $\boldsymbol{y}_1^{(j)}$ of MVLR j is

$$\begin{aligned}\boldsymbol{y}_1^{(j)} &= \boldsymbol{d}_1^{(j)} + \varepsilon_1^{(j)} = \phi(\boldsymbol{C}^{(j)}) + \varepsilon_1^{(j)} = \log_{10}(\boldsymbol{C}^{(j)} + 0.1) + \varepsilon_1^{(j)} \\ &= \boldsymbol{X}^{(j)}\boldsymbol{\alpha}^{(j)} + \alpha_0^{(j)}\boldsymbol{1} + \varepsilon_1^{(j)} = \tilde{\boldsymbol{X}}^{(j)}\tilde{\boldsymbol{\alpha}}^{(j)} + \varepsilon_1^{(j)}\end{aligned} \tag{4}$$

Here, $\boldsymbol{X}^{(j)} \in R^{N_j\times m}$, $\boldsymbol{d}_1^{(j)} = \log_{10}(\boldsymbol{C}^{(j)} + 0.1) \in R^{N_j\times 1}$, $\boldsymbol{\alpha}^{(j)} \in R^m$ is a coefficient vector, $\alpha_0^{(j)}$ a constant, $\boldsymbol{1}$ a vector of

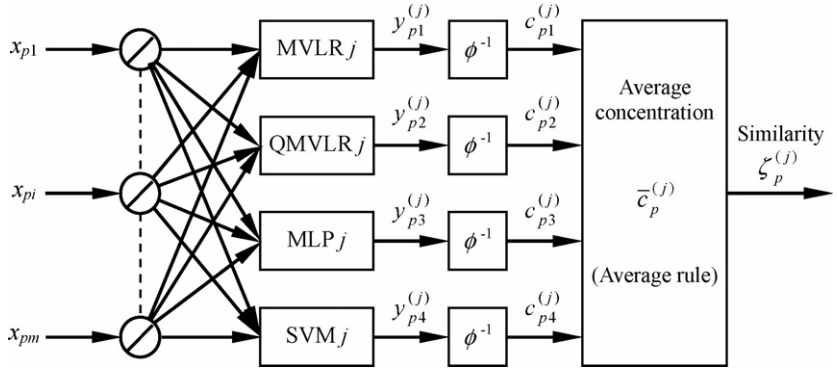

Fig. 4. Block diagram of approximation model ensemble j.

length N_j consisting solely of 1's, $\tilde{\boldsymbol{X}}^{(j)} = \{\boldsymbol{X}^{(j)}, \mathbf{1}\} \in R^{N_j \times (m+1)}$, called the augmented matrix of $\boldsymbol{X}^{(j)}$, and $\tilde{\boldsymbol{\alpha}}^{(j)} = \{\boldsymbol{\alpha}^{(j)}, \alpha_0^{(j)}\}$. Obviously, there are $m+1$ undecided parameters. The $m+1$-dimensional parameter vector $\tilde{\boldsymbol{\alpha}}^{(j)}$ is determined by the least-mean-squared (LMS) method

$$\tilde{\boldsymbol{\alpha}}^{(j)}(0) = 1,$$

$$\tilde{\boldsymbol{\alpha}}^{(j)}(\tau+1) = \tilde{\boldsymbol{\alpha}}^{(j)}(\tau) + \eta(\tilde{\boldsymbol{X}}^{(j)}(\boldsymbol{d}_1^{(j)} - \tilde{\boldsymbol{X}}^{(j)}\tilde{\boldsymbol{\alpha}}^{(j)}(\tau)))^{\mathrm{T}} \quad (5)$$

where $\eta > 0$ is a learning rate.

In order to make it stand out clearly, 10^x is marked as pow10(x), or pow10(x) = 10^x. The predicted concentration of $\boldsymbol{x}_p$ given by MVLR j is

$$c_{p1}^{(j)} = \phi^{-1}(y_{p1}^{(j)}) = \mathrm{pow10}(y_{p1}^{(j)}) - 0.1$$
$$= \mathrm{pow10}((\boldsymbol{\alpha}^{(j)})^{\mathrm{T}}\boldsymbol{x}_p + \alpha_0^{(j)}) - 0.1 \quad (6)$$

3.3. Quadratic multivariate logarithmic regression models

The learning process of QMVLR j is to realize the mapping $f : \boldsymbol{X}^{(j)} \to \boldsymbol{d}_2^{(j)}$ and minimize the RMS error $\varepsilon_2^{(j)} = \sqrt{1/2N_j}||\boldsymbol{d}_2^{(j)} - \boldsymbol{y}_2^{(j)}||_2$. Here, $\boldsymbol{y}_2^{(j)}$ the real output vector of QMVLR j, and $\boldsymbol{d}_2^{(j)} = \boldsymbol{d}_1^{(j)}$. For the pth training sample $\boldsymbol{x}_p^{(j)} = (x_{p1}^{(j)}, \ldots, x_{pi}^{(j)}, \ldots, x_{pm}^{(j)})^{\mathrm{T}} \in R^m$, the real output of QMVLR j is

$$y_{p2}^{(j)} = d_{p2}^{(j)} + \varepsilon_{p2}^{(j)} = \log_{10}(C_p^{(j)} + 0.1) + \varepsilon_{p2}^{(j)}$$
$$= \sum_{i=1}^{m} A_i^{(j)}(x_{pi}^{(j)})^2 + \sum_{i=1}^{m}\beta_i^{(j)}x_{pi}^{(j)} + \beta_0^{(j)} + \varepsilon_{p2}^{(j)} \quad (7)$$

In order to simplify the computation, the cross terms, $x_{pi}^{(j)}x_{pk}^{(j)}$, are emitted. Let $\tilde{\boldsymbol{X}}^{(j)} = \{(\boldsymbol{X}^{(j)})^2, \boldsymbol{X}^{(j)}, \mathbf{1}\} \in R^{N_j \times (2m+1)}$. The $2m+1$ undecided parameters are also determined with the same LMS learning method as used in (5).

The predicted concentration of $\boldsymbol{x}_p$ given by QMVLR j is

$$c_{p2}^{(j)} = \phi^{-1}(y_{p2}^{(j)}) = \mathrm{pow10}(y_{p2}^{(j)}) - 0.1$$
$$= \mathrm{pow10}\left(\sum_{i=1}^{m} A_i^{(j)}x_{pi}^2 + \sum_{i=1}^{m}\beta_i^{(j)}x_{pi} + \beta_0^{(j)}\right) - 0.1 \quad (8)$$

3.4. Single-hidden-layer perceptrons

Let the structure of MLP j in ensemble j be m-s-1, which works for the many-to-one mapping $f : \boldsymbol{X}^{(j)} \to \boldsymbol{d}_3^{(j)}$. Here, s is the number of hidden nodes. Every MLP uses the batch-learning back-propagation algorithm to adjust its weights and biases. The learning process of MLP j is to minimize the RMS error $\varepsilon_3^{(j)} = \sqrt{1/2N_j}||\boldsymbol{d}_3^{(j)} - \boldsymbol{y}_3^{(j)}||_2$ between $\boldsymbol{d}_3^{(j)}$ and the real output $\boldsymbol{y}_3^{(j)}$.

It is shown in theory that variable scales and activation functions have great influences on the generalizations and convergences of MLPs [19,23,24]. In our experiment, the activation functions in the hidden and the output layers are $f(\varphi) = 3(1 + \exp(-\varphi/3))^{-1}$, all the input variables scaled in proportion to the range of [0.0, 6.0], and the target outputs to the range of [0.0, 3.0] [23]. The real output of MLP j is written as

$$y_{p3}^{(j)} = f\left(\sum_{h=0}^{s} w_{jh} f\left(\sum_{i=0}^{m} w_{hi}^{(j)} x_{pi}\right)\right) \quad (9)$$

Here, $w_{hi}^{(j)}$ is the weight component between input node i and hidden unit h, and w_{jh} the weight component between h and output node j.

Let the real concentration of $\boldsymbol{x}_p$ be $C_p^{(j)}$ (ppm). The target output $d_{p3}^{(j)}$ of MLP j is given by

$$d_{p3}^{(j)} = (1 + \log_{10}(C_p^{(j)} + 0.1))\frac{3.0}{5.0} \quad (10)$$

In consideration of the cut-off and saturated states of sigmoid activation functions, the range of target outputs should be further cut down. Let $d_{p3}^{(j)} = 0.1$ correspond to $C_p^{(j)} = 0$ ppm (distilled water) and $d_{p3}^{(j)} = 2.9$ to $C_p^{(j)} = 10,000$ ppm. Then (11) is revised to be:

$$d_{p3}^{(j)} = \phi(C_p^{(j)}) = (1 + \log_{10}(C_p^{(j)} + 0.1))\frac{2.8}{5.0} + 0.1 \quad (11)$$

The predicted concentration of $\boldsymbol{x}_p$ given by MLP j is:

$$c_{p3}^{(j)} = \phi^{-1}(y_{p3}^{(j)}) = \mathrm{pow10}\left(\frac{5.0(y_{p3}^{(j)} - 0.1)}{2.8} - 1\right) - 0.1 \quad (12)$$

3.5. Support vector machines

SVM j realizes the many-to-one mapping $f : \boldsymbol{X}^{(j)} \to \boldsymbol{d}_4^{(j)}$ through learning, and $\boldsymbol{d}_4^{(j)} = \boldsymbol{d}_1^{(j)}$. The learning process of SVM j is evaluated by the RMS error $\varepsilon_4^{(j)} = \sqrt{1/2N_j}||\boldsymbol{d}_4^{(j)} - \boldsymbol{y}_4^{(j)}||_2$. We use the RBF kernels $k(\boldsymbol{x}, \boldsymbol{x}_q) = \exp(-\gamma||\boldsymbol{x} - \boldsymbol{x}_q||_2^2)$ in every SVM, where γ is a width parameter determined artificially, and $\boldsymbol{x}_q$ a support vector. The real output of SVM j for $\boldsymbol{x}_p$ is

$$y_{p4}^{(j)} = f(\boldsymbol{x}_p, \boldsymbol{X}_{\mathrm{sup}}^{(j)}, \boldsymbol{b}^{(j)}) = \sum_{q=1}^{\tilde{N}_j} b_q^{(j)} k(\boldsymbol{x}_p, \boldsymbol{x}_q^{(j)}) + b_0^{(j)} \quad (13)$$

Here, $\boldsymbol{X}_{\mathrm{sup}}^{(j)}$ is the jth support vector matrix, $\tilde{N}_j$ the number of support vectors, $b_q^{(j)}$ the weight of the qth kernel related to $\boldsymbol{x}_q^{(j)}$, and $b_0^{(j)}$ a constant. All of them are determined by solving the following optimization problem in the primal weight space

$$\min_{b,\xi,\xi^*} J(\boldsymbol{b}, \boldsymbol{\xi}, \boldsymbol{\xi}^*) = \frac{1}{2}\sum_{p=1}^{N_j}(b_p^{(j)})^2 + C\sum_{p=1}^{N_j}(\xi_p^{(j)} + \xi_p^{(j)*})$$

$$\text{such that} \begin{cases} d_{p4}^{(j)} - y_{p4}^{(j)} \le \varepsilon + \xi_p^{(j)} \\ y_{p4}^{(j)} - d_{p4}^{(j)} \le \varepsilon + \xi_p^{(j)*} \\ \xi_p^{(j)}, \xi_p^{(j)*} \ge 0 \end{cases} \quad (14)$$

where ε is an insensitivity constant and C a capacity constant determined in prior. In addition, $\xi_p^{(j)}$ and $\xi_p^{(j)*}$ are slack vari-

ables, determined through learning. The detailed algorithm can be found in [21].

The predicted concentration of $\boldsymbol{x}_p$ given by SVM j is

$$c_{p4}^{(j)} = \phi^{-1}(y_{p4}^{(j)}) = \text{pow10}(y_{p4}^{(j)}) - 0.1$$
$$= \text{pow10}\left(\sum_{q=1}^{\tilde{N}_j} b_q^{(j)} k(\boldsymbol{x}_p, \boldsymbol{x}_q^{(j)}) + b_0^{(j)}\right) - 0.1 \quad (15)$$

4. Experimental results

Four kinds of fragrant materials are ethanol, ethyl acetate, ethyl caproate and ethyl lactate, and each of them is diluted with distilled water into 4–6 kinds of concentrations. The measurement method is just the same as stated in Section 2. Fifty samples for each concentration are collected, 40 of them are used as a part of the training set, and the remains as a part of the test set. As a result, there are $40 \times 23 = 920$ samples in the training set, and $10 \times 23 = 230$ patterns in the test set. Fig. 5 is the principle component analysis (PCA) result for all the original dataset [9]. The reproducibility of data is relatively good. Obviously the four many-to-one curves are both non-linear and different from each other.

There are four ensembles for four kinds of odors, corresponding to ethanol, ethyl acetate, ethyl caproate and ethyl lactate, and four experts in each, namely an MVLR, a QMVLR, an MLP, and an SVM. Simply, the four ensembles are respectively indexed as "e", "ea", "ec" and "el" hereafter. The number of input dimensions is 16, because there are 16 sensors in the array. Since the measurements of distilled water are taken for the original points of curves, the numbers of samples are 280 in the training subsets "e", "ec" and "el", and 200 in the training subset "ea".

Let the learning rate be $\eta = 0.10$, and the maximum number of iteration steps $\tau_{\max} = 1000$. The MVLR and QMVLR models are realized with MATLAB 6.2. The learning time of each model is within 1 s (P2.6G 256 RAM, the same below). As a result, the RMS errors of predicted logarithmic concentrations of 70 testing ethyl caproate samples given by MVLR "ec" and QMVLR "ec" are 0.084 and 0.065, respectively.

For the MLP experts, let the number of hidden nodes be $s = 5$, the learning factor $\eta = 0.02$, the momentum parameter $\alpha = 0.075$, and the maximum iteration epoch $\tau_{\max} = 15{,}000$. Perceptrons "e", "ec" and "el" take 7.11 s, and perceptron "ea" 5.19 s. The RMS errors of predicted logarithmic concentrations of the test samples given by the MLP experts for the represented odors are 0.025, 0.027, 0.030, and 0.019 in order. Such results are satisfactory. A 16-5-1 perceptron only needs to store $(17 \times 5 + 6 \times 1)/16 = 5.69$ vectors.

Let the width parameter $\gamma = 1/8$, the insensitivity constant $\varepsilon = 0.05$, and the capacity constant $C = 1000$. All the learning time of the four SVM experts is within 5 s. The numbers of support vectors in each SVM are 30, 72, 91, and 52 in sequence. As an example, the RMS error of predicted logarithmic concentrations of 70 ethyl caproate testing samples by SVM expert "ec" is 0.023.

Table 2
R.S.D.s of predicted concentrations of odors given by four four-membered ensembles (%)

Ensemble	Odor																				
	Ethanol concentration (ppm)						Ethyl acetate concentration (ppm)				Ethyl caproate concentration (ppm)					Ethyl lactate concentration (ppm)					
	100	1000	2500	5000	7500	10000	10	100	1000	10000	1	5	10	100	1000	10000	100	1000	2500	5000	7500	10000
1	17.63	26.85	35.93	31.45	26.40	27.30	138.81	181.80	35.78	57.28	38.01	45.98	118.23	126.68	198.20	199.25	53.43	151.00	187.90	193.56	183.65	141.25
2	143.41	61.60	167.56	189.45	194.16	194.27	22.90	17.45	35.95	20.71	39.81	109.58	191.11	178.69	53.84	32.65	168.52	127.38	7.99	60.85	101.73	151.95
3	96.70	200.00	200.00	200.00	200.00	200.00	35.19	99.38	200.00	200.00	24.35	21.42	24.61	27.18	20.90	18.61	27.39	136.45	97.93	199.27	200.00	200.00
4	105.35	148.79	71.34	89.45	105.13	128.69	43.04	56.63	199.84	199.99	36.82	20.78	73.24	132.40	195.81	194.19	21.51	26.62	24.52	24.57	31.29	31.55

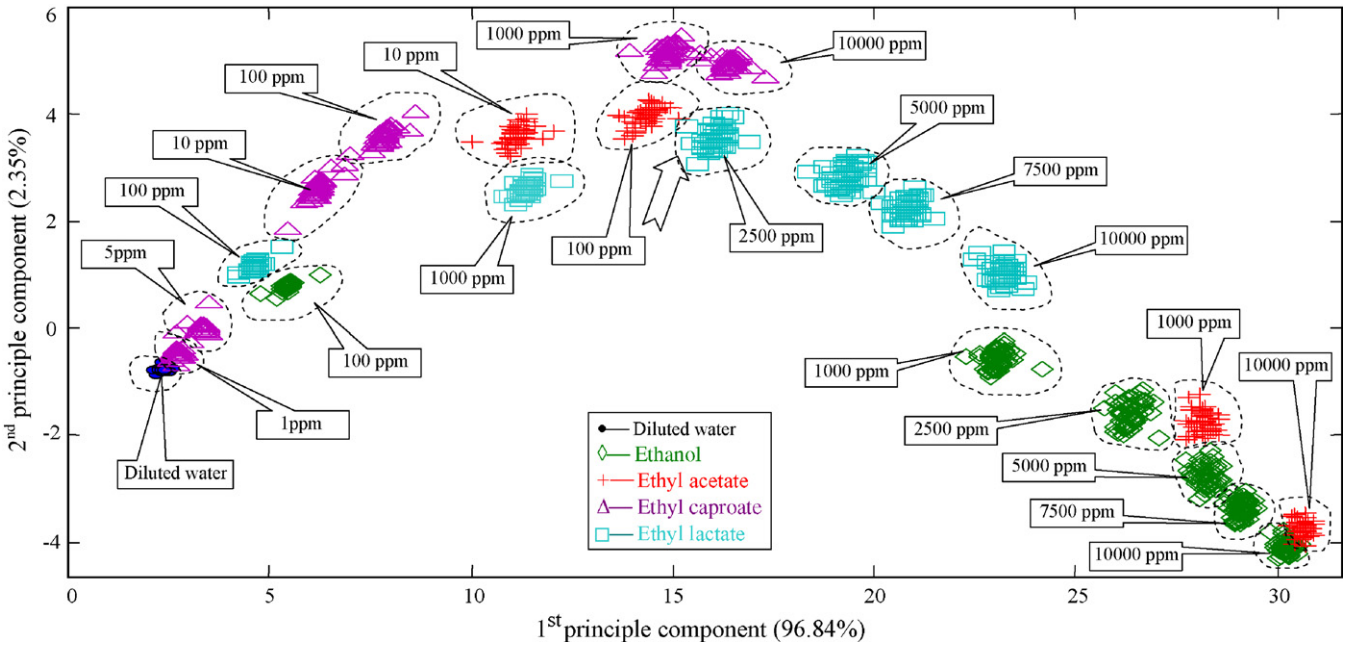

Fig. 5. Principle component analysis for the original data set.

Relatively speaking, the approximation performance of an SVM model is the best among four experts. From the angle of storage requirement, however, an SVM expert does not occupy any dominant position. An SVM often needs to store too many support vectors, compared with an MVLR, a QMVLR, or an MLP expert.

It has been made clear in Section 3.1 that we cannot depend upon any single type of approximation models to determine

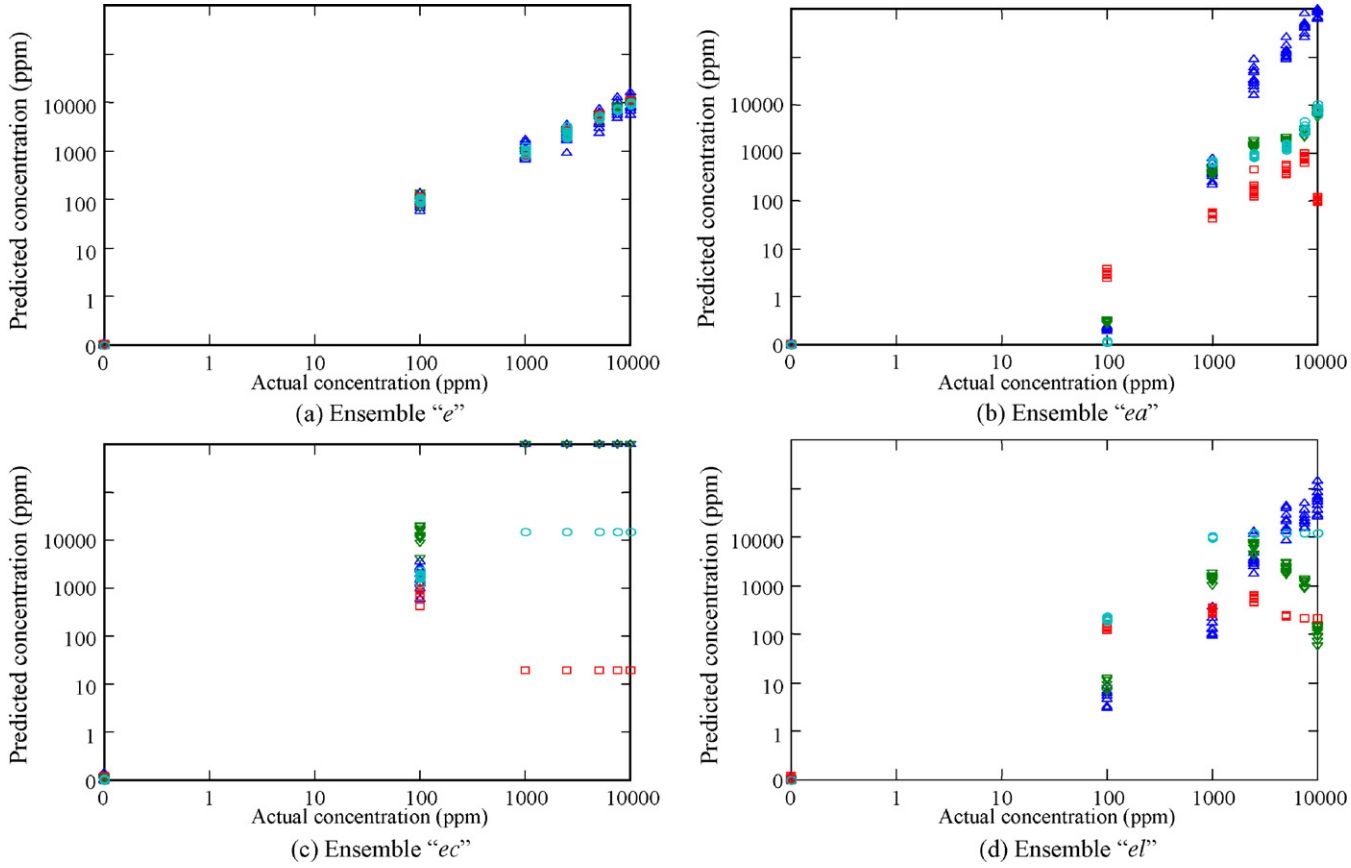

Fig. 6. Predictions of ethanol by the four-membered ensembles (logarithmic coordinates). (△) MVLR; (▽) QMVLR; (◯) MLP; (□)-SVM.

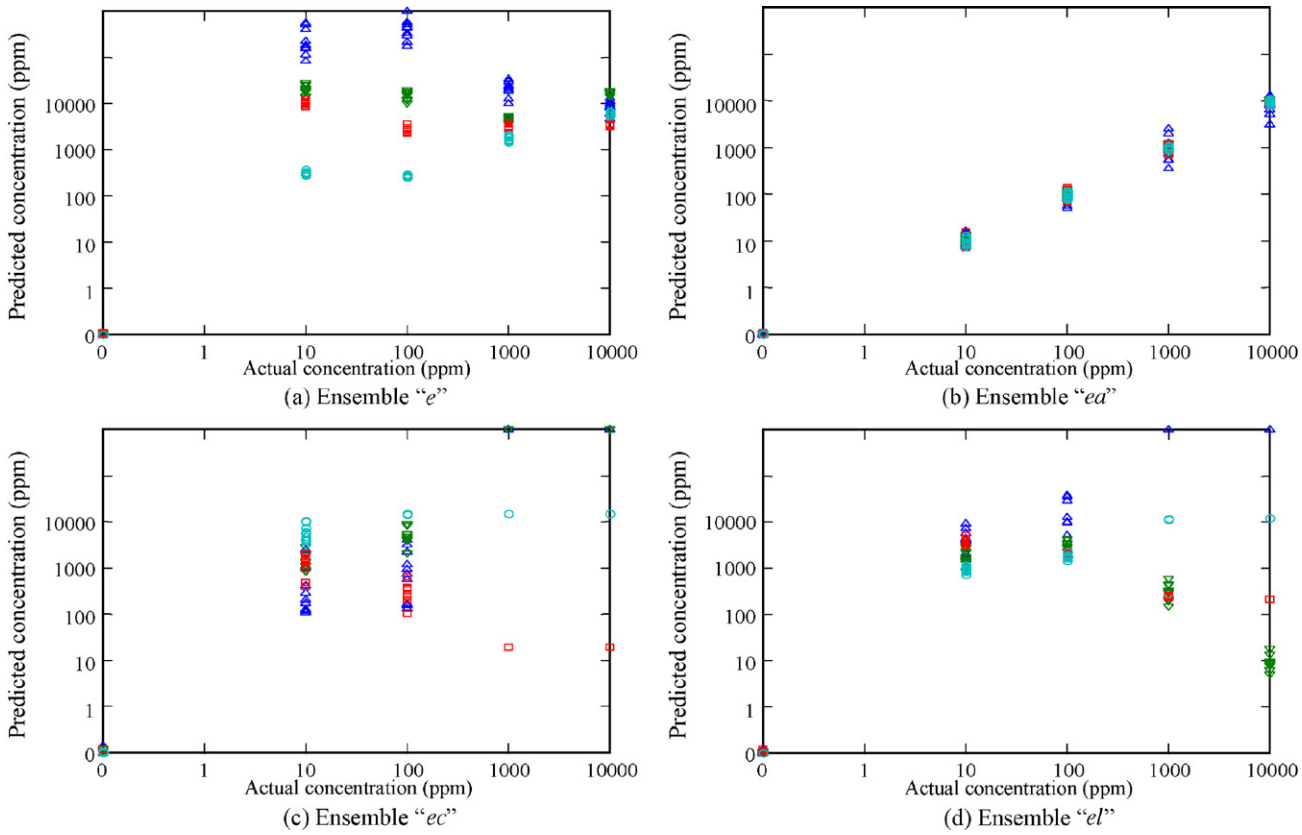

Fig. 7. Predictions of ethyl acetate by the four-membered ensembles (logarithmic coordinates). (△) MVLR; (▽) QMVLR; (◯) MLP; (□)-SVM.

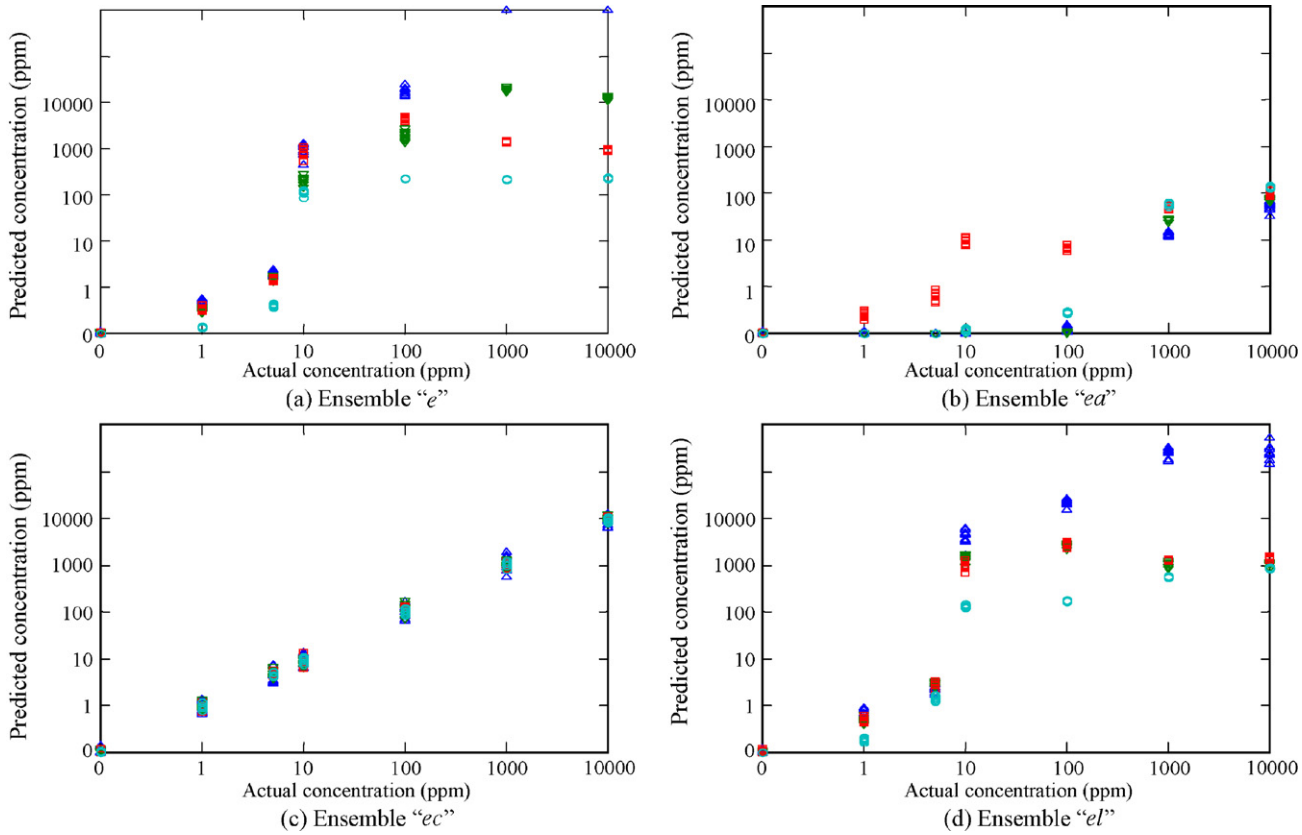

Fig. 8. Predictions of ethyl caproate by the four-membered ensembles (logarithmic coordinates). (△) MVLR; (▽) QMVLR; (◯) MLP; (□)-SVM.

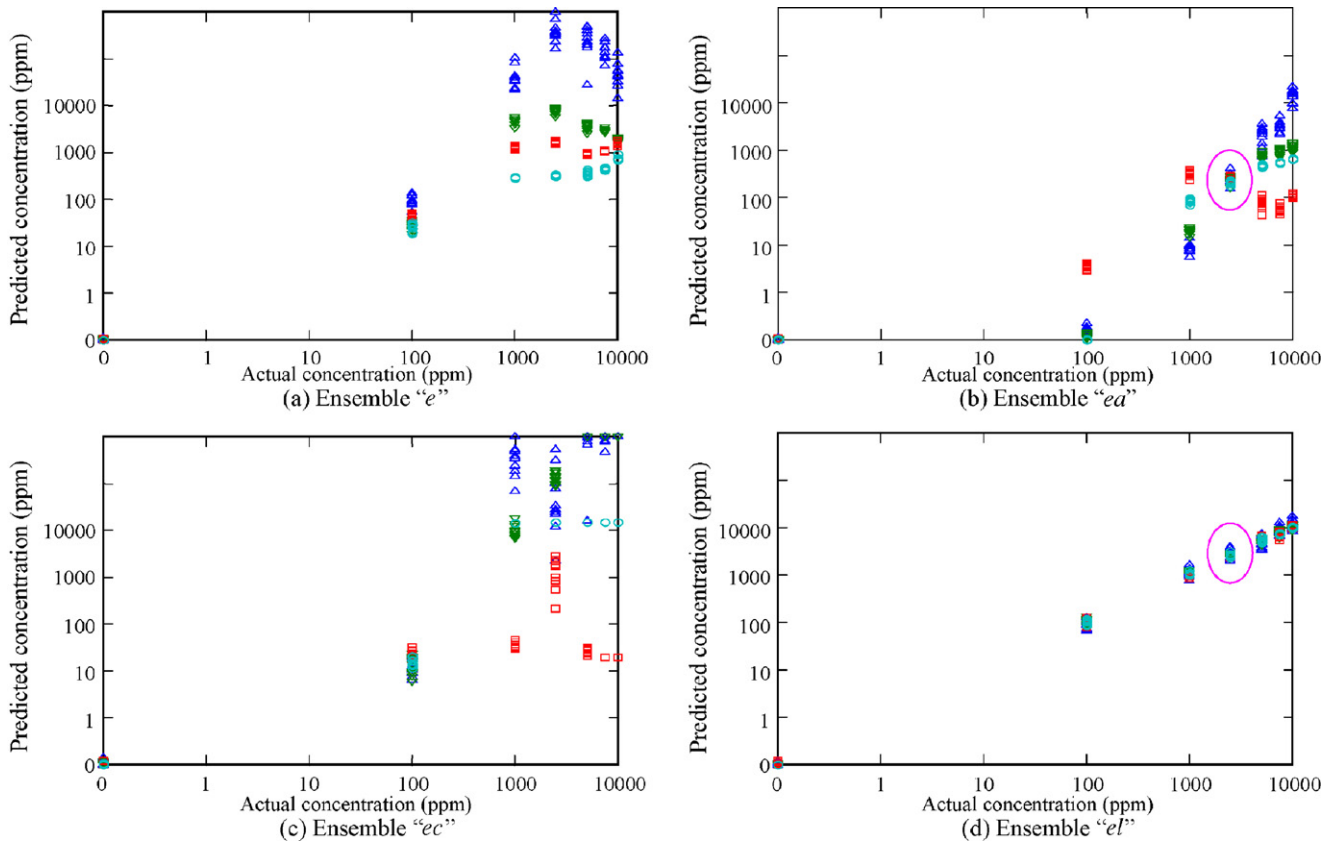

Fig. 9. Predictions of ethyl lactate by the four-membered ensembles (logarithmic coordinates). (△) MVLR; (▽) QMVLR; (◯) MLP; (□)-SVM.

the labels and concentrations of odor samples. Therefore, the predicted results given by a single kind of many-to-one approximation models are not worthy convincing. Figs. 6–9 give the predicted results of the test set given by the four four-membered ensembles for their represented and non-represented odors. In order to make the figures clearly, the following restrictions are exerted. If a predicted concentration given by an expert is over 10^6 ppm, it is forced to be 10^6 ppm; and if below 0.1 ppm, to be equal to 0.1 ppm. The following phenomena are quite obvious. For a represented odor, the predicted concentrations of samples given by four experts in its representative ensemble are quite close to each other, otherwise quite divergent. For example, a 10,000 ppm ethyl acetate is unexpectedly taken for a $10^{19.32}$ ppm ethyl caproate by MVLR "ec" or a $10^{80.40}$ ppm ethyl caproate by QMVLR "ec", as shown in Fig. 7(c). How absurd that phenomenon is! Conversely, a 10 ppm ethyl caproate is looked upon as a $10^{-1.28}$ ppm ethyl acetate by MVLR "ea" or $10^{-1.40}$ ppm ethyl acetate by QMVLR "ea", as shown in Fig. 8(b). In addition, the scales of predicted outputs given by an MLP or an SVM expert are between 0.1 and 10,000 ppm. Here, we notice with gratification that MVLR or QMVLR models are not very good in function approximation capabilities, but they play a unique role in saying "No". These differences are just our basis of decisions making.

Table 2 gives the R.S.D.s of predicted concentrations given by the four four-membered ensembles for four kinds of odors. The values in boldface are the maximum R.S.D.s given by the representative panels, otherwise the minimum given by the non-representative. Table 2 proves the following facts again that the R.S.D.s of predicted concentrations of samples given by the representative ensembles are relatively small; otherwise very large. In this way, we can simultaneously determine many kinds of odor classes and concentrations. According to Figs. 6–9, a solely confusable place is that some 2500 ppm ethyl lactate sam-

Table 3
An example of why a 2500-ppm ethyl lactate sample is mislabeled as a 212-ppm ethyl acetate sample (ppm)

Ensemble	MVLR	QMVLR	SVM	MLP	Average concentration	Standard deviation	R.S.D. (%)
e	336884.00	6898.74	3745.34	1333.31	86465.3	166969	193.11
ea	207.17	201.67	237.58	203.07	212.37	16.97	7.99
ec	103362.00	191845.00	2392.93	14838.00	78109.40	88147.00	112.85
el	1989.94	2583.59	2727.49	2661.12	2490.54	338.87	13.61

ples are probably taken for about 200 ppm ethyl acetate ones, marked by circles in Fig. 9(b) and (d). In line with Table 2, two 2500 ppm ethyl lactate samples are finally mislabeled as 210 ppm ethyl acetate ones. For the two misclassified samples, the average predicted concentrations given by ensemble "el" are 2490 and 2465 ppm, and the predicted R.S.D.s 13.61% and 24.52%. The corresponding values given by ensemble "ea" are 212 and 171 ppm, and the predicted R.S.D.s 7.99% and 12.49%. Table 3 gives a detailed example of why a 2500 ppm ethyl lactate sample is labeled as a 212 ppm ethyl acetate sample. The reason is that the R.S.D. of average predicted concentration given by the latter expert ensemble is smaller than that given by the former. In fact, the PCA result also tells us the existence of this kind of possibility, as shown with an arrow in Fig. 5. As a result, the predicted correct rate of the test set given by the proposed approximation model ensembles is 228/230 = 99.13%. Such results are quite satisfactory.

5. Conclusions

This paper sets up a practical electronic nose. After decomposing a many-to-many approximation task into multiple many-to-one approximation ones, we employ MVLRs, QMVLRs, MLPs and SVMs as the basic components of ensembles to implement the simultaneous estimation tasks of odor classes and concentrations. Each member of an ensemble only needs to be trained by samples from its represented odor, which only account for a small part of the original training dataset. And this paper uses the average and the minimum combination rules to determine the final labels and concentrations of odor samples. The experimental results for estimating four kinds of fragrant materials, 21 grades of concentrations in total, show that the proposed approximation model ensembles and combination strategies are quite effective, so improves the recognition performances of electronic noses. With the proposed function approximation model ensembles and combination strategies, an electronic nose can well execute the simultaneous estimation tasks of many kinds of odor classes and concentrations.

Acknowledgments

This work is supported by the National Science Foundation of China (NSFC) under Grant no. 60275017, 60373073; the Key Science and Technology Development Foundation of Shanghai, China, under Grant no. 025115028, 04*dz*05010; and the open project program of the State Key Laboratory of Bioreactor Engineering.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.snb.2006.03.017.

References

[1] T.C. Pearce, S.S. Schiffman, H.T. Nagle, J.W. Gardner (Eds.), Hand of Machine Olfaction, Wiley-VCH Press, 2003.
[2] J.W. Gardner, P.N. Bartlett, Electronic Noses: Principles and Applications, Oxford University Press, 1999.
[3] M.S. Nayak, R. Dwivedi, S.K. Srivastava, Application of iteration technique in association with multiple regression method for identification of mixtures of gases using an integrated gas-sensor array, Sens. Actuators B: Chem. 21 (1994) 11–16.
[4] S. Capone, M. Epifani, F. Quaranta, P. Siciliano, A. Taurino, L. Vasanelli, Monitoring of rancidity of milk by means of an electronic nose and a dynamic PCA analysis, Sens. Actuators B: Chem. 78 (2001) 174–179.
[5] G. Huyberechts, P. Szecowka, J. Roggen, B.W. Licznerski, Simultaneous quantification of carbon monoxide and methane in humid air using a sensor array and an artificial neural network, Sens. Actuators B: Chem. 45 (1997) 123–130.
[6] A.K. Srivastava, Detection of volatile organic compounds (VOCs) using SnO_2 gas-sensor array and artificial neural network, Sens. Actuators B: Chem. 96 (2003) 24–37.
[7] Z. Haoxian, M.O. Balaban, J.C. Principe, Improving pattern recognition of electronic nose data with time-delay neural networks, Sens. Actuators B: Chem. 96 (2003) 385–389.
[8] G. Daqi, W. Shuyan, J. Yan, An electronic nose and modular radial basis function network classifiers for recognizing multiple fragrant materials, Sens. Actuators B: Chem. 97 (2004) 391–401.
[9] G. Daqi, C. Mingming, J. Yan, Simultaneous estimation of classes and concentrations of odors by an electronic nose using combinative and modular multilayer perceptrons, Sens. Actuators B: Chem. 107 (2005) 773–781.
[10] M. Vinaixa, E. Llobet, J. Brezmes, X. Vilanova, X. Correig, A fuzzy ARTMAP- and PLS-based MS e-nose for the qualitative and quantitative assessment of rancidity in crisps, Sens. Actuators B: Chem. 106 (2005) 677–686.
[11] A. Szczurek, M. Maciejewska, Relationship between odor intensity assessed by human assessor and TGS sensor array response, Sens. Actuators B: Chem. 106 (2005) 13–19.
[12] L.K. Hansen, P. Salamon, Neural network ensembles, IEEE Trans. Pattern Anal. Mach. Intell. 12 (1990) 993–1001.
[13] J. Kittier, M. Hatef, R.P.W. Duin, J. Matas, On combining classifiers, IEEE Trans. Pattern Anal. Mach. Intell. 20 (1998) 226–239.
[14] D. Windridge, J. Kittler, A morphologically optimal strategy for classifier combination: multiple expert fusions as a tomographic process, IEEE Trans. Pattern Anal. Mach. Intell. 25 (2003) 343–353.
[15] C.A. Shipp, L.I. Kuncheva, Relationships between combination methods and measures of diversity in combining classifiers, Inform. Fusion 3 (2002) 135–148.
[16] G. Uda, M. Amel, Modular neural network classifiers: a comparative study, J. Intell. Robotic Syst. 21 (1998) 117–129.
[17] R. Anand, K.G. Mehrotra, C.K. Mohan, S. Ranka, Efficient classification for multiclass problems using modular neural networks, IEEE Trans. Neural Networks 6 (1995) 117–124.
[18] R.O. Duda, P.E. Hart, D.G. Stork, Pattern Classification, 2nd ed., John Wiley and Sons Inc., 2001.
[19] C.M. Bishop, Neural Networks for Pattern Recognition, Clarendon Press, Oxford, 1995.
[20] V.N. Vapnik, Statistical Learning Theory, Wiley, New York, 1998.
[21] J.A.K. Suykens, T.V. Gestel, J.D. Brabanter, B.D. Moor, J. Vandewalle, Least Squares Support Vector Machines, World Scientific Publishing Company Private Limited, 2002.
[22] C. Distante, N. Ancona, P. Siciliano, Support vector machines for olfactory signals recognition, Sens. Actuators B: Chem. 88 (2003) 30–39.
[23] G. Daqi, Y. Genxing, Influences of variable scales and activation functions on the performances of multilayer feedforward neural networks, Pattern Recogn. 36 (2003) 869–878.
[24] G. Thimm, P. Moerland, E. Fiesler, The interchangeability of learning rate and gain in backpropagation neural networks, Neural Comput. 8 (1996) 451–460.

Biographies

Gao Daqi received his PhD degree from Zhejiang University, China, in 1996. Currently, he is a professor in East China University of Science and Technology. His research interests are pattern recognition, neural networks, and machine olfactory.

Chen Wei is currently a Master student in East China University of Science and Technology. His research interests include neural networks and pattern recognition.