Application of the new pattern recognition system in the new e-nose to detecting Chinese spirits*

Gu Yu(谷 宇)† and Li Qiang(李 强)‡

Department of Mechanics, School of Civil Engineering, Beijing Jiaotong University, Beijing 100044, *China*

(Received 25 November 2013; revised manuscript received 31 December 2013; published online 20 February 2014)

We present a new pattern recognition system based on moving average and linear discriminant analysis (LDA), which can be used to process the original signal of the new polymer quartz piezoelectric crystal air-sensitive sensor system we designed, called the new e-nose. Using the new e-nose, we obtain the template datum of Chinese spirits via a new pattern recognition system. To verify the effectiveness of the new pattern recognition system, we select three kinds of Chinese spirits to test, our results confirm that the new pattern recognition system can perfectly identify and distinguish between the Chinese spirits.

Keywords: new pattern recognition system, new e-nose, detecting, Chinese spirits

PACS: 42.30.Sy, 43.60.Lq, 07.07.Df **DOI:** 10.1088/1674-1056/23/4/044213

1. Introduction

The pattern recognition system is an important part of both information science and artificial intelligence and is widely used in intelligent robots to provide their required intelligence, information, networking, and computing. Research on, and breakthroughs in, the pattern recognition system for intelligent robots has hitherto primarily focused on visual, auditory, and tactile aspects, while the use of the pattern recognition system for applications of taste and smell remains rare. There are two reasons for this disparity. Primarily, researchers cannot obtain highly sensitive taste and smell sensors. Secondly, the sensors must be assembled with highly developed pattern recognition systems. The two aforementioned aspects heavily influence the application of the e-nose. Researchers have already done much research in an attempt to resolve these problems. Gu *et al.*[1] investigated the structure and absorbability of nanometer nonmetallic polymer thin film coating on quartz, and also performed the vibration analysis of new polymer quartz piezoelectric crystal sensor for detecting characteristic materials of volatility liquid. Muhammad *et al.*[2] studied the orange dye polyaniline composite-based impedance humidity sensors; Sekhar *et al.*[3] used electrochemical potentiometric gas sensors to trace the detection and discrimination of explosives. Ruys *et al.*[4] used a coated piezoelectric sensor to detect mercury in air. Ng *et al.*[5] explored flexible image sensor array with bulk heterojunction organic photodiode. Steibig[6] reported on a three-color sensor array based on a vertically integrated thin-film structure. Guo *et al.*[7] provided an LDA-based sensor selection approach used in breath analysis. Potyrailo and Morris[8] did research on a wireless resonant sensor array for high-throughput screening of materials. Kildishev *et al.*[9] studied the magnetic field beyond a rectangular array of sensors. Gao *et al.*[10] reported on the multivariate recurrence network for characterizing horizontal oil–water two-phase flow. The pattern recognition system is still in the exploration and rapid development stage, with new methods and new theories emerging constantly. The neural network algorithm, genetic algorithm, fuzzy reasoning, and optimal control theory are the main methods currently used. We designed a new e-nose for detecting Chinese spirits. The new e-nose contains a sensor array, a pneumatic transmission path and a recognition system. An effective pattern recognition system is the most important part of the new e-nose, but the existing methods cannot effectively extract the feature vector and explore useful information maximums. We present a new pattern recognition system based on the moving average and linear discriminant analysis (LDA) that is adopted in the new e-nose for detecting Chinese spirits. The verification of the effectiveness of the new pattern recognition system is carried out using the new e-nose to detect 3 kinds of Chinese spirits.

2. Theory and application

The new pattern recognition system is based on moving average and linear discriminant analysis (LDA).

The dynamic test data $y(t)$ consists of the determined element $f(t)$ and the random element $e(t)$. The determined element $f(t)$ is the characteristic information we need, while the random element $x(t)$ is the random fluctuant error of testing. After disposing the signal of the sample, the dynamic test data

*Project supported by the National High Technology Research and Development Program of China (Grant No. 2013AA030901).

†Corresponding author. E-mail: yugu@bjtu.edu.cn,

‡E-mail: 12121028@bjtu.edu.cn

© 2014 Chinese Physical Society and IOP Publishing Ltd

can be denoted as

$$y_i(t) = f_i(t) + e_j(t), \quad i, j = 1, 2, \ldots, N. \tag{1}$$

In order to describe the test results accurately, we must restrain the effects of random error (e_j). We always dispose the dynamic test data $\{y_j\}$ through the moving average and smoothing process. More specifically, in an appropriate section, we process the non-stationary datum $\{y_j\}$ into stable datum in order to reduce the random fluctuation of the random error (e_j). With the moving average, we can obtain a relatively smooth test result (f_j) and filter out the random fluctuation error e_j.

The principle of LDA is as follows. First, for the advanced copy x, we assume the linear discriminant function $g(x)$, and we can, according to problem C, define the corresponding discriminant function. Secondly, we substitute the unknown sample X into the discriminant function wherein the number is C, and fit that sample into the problem of the maximum discriminant function. The core of the LDA is to define the weight vector and threshold power of the function according to the known sample.

Between-class scatter matrix of the sample can be defined as

$$S_{\mathrm{b}} = \sum_{i=1}^{C} P_i(\mu_i - \mu)^{\mathrm{T}}, \tag{2}$$

where μ_i is the average value of category C_i, μ is the average value of all samples, P_i is the prior probability.

Within-class scatter matrix of the sample can be defined as

$$S_{\mathrm{w}} = \sum_{i=1}^{C} P_i S_i, \tag{3}$$

where

$$S_i = E[(x - \mu_i)(x - \mu_i)^{\mathrm{T}} | x \in C_i].$$

In order to extract more useful information, we hope that small within-class scatter and large between-class scatter of simples occurs after the samples have been projected. Therefore, if S_{w} is a nonsingular matrix, the best projection direction W_{opt} are orthogonal eigenvectors, which makes the value of the determinantal ratio between between-class scatter matrix and within-class scatter matrix reach the maximum. So the equation of the relevant Fisher's linear discriminant can be defined as

$$J(W_{\mathrm{opt}}) = \arg\max \frac{|W^{\mathrm{T}} S_{\mathrm{b}} W|}{|W^{\mathrm{T}} S_{\mathrm{W}} W|}, \tag{4}$$

where W_{opt} is the value that satisfies the following equation:

$$S_{\mathrm{b}} W_i = \lambda_i S_{\mathrm{w}} W_i \quad (i = 1, 2, \ldots, m). \tag{5}$$

That is to say, specific to the relative huge eigenvector of the eigenvalue of matrix $S_{\mathrm{w}}^{-1} S_{\mathrm{b}}$, this matrix has a number of $C-1$ nonzero eigenvalues at most, where C is the value of the category.

The new pattern recognition system can be used in the new e-nose, the new polymer quartz piezoelectric crystal air-sensitive sensor system, which contains a sensing unit, a pneumatic transmission path, and a back-end circuit. The sensing unit can discern the measured information, the pneumatic path can guide gas in flowing, and the back-end circuit can capture the signal and display detecting solutions.

More specifically, the new sensor takes quartz piezoelectric crystal as the basal material and has a nanometer nonmetallic polymer thin film coating on its surface. The coating, plated by vacuum electron beam dispersion (EBD) coating technology, is carried out by employing different nanometer nonmetallic polymer thin films that have properties of selecting and identifying characteristic materials. Due to the physical absorption of the polymer to the characteristic substance of liquor, the system can circulate. The processing of the system can be divided into the cleaning and absorption stages, and the back-end circuit can monitor and record the frequency signals of the sensor system. As shown in Figs. 1 and 2, the signal distribution of the sensor can be found, where, x represents time in seconds, and y represents the frequency in Hz.

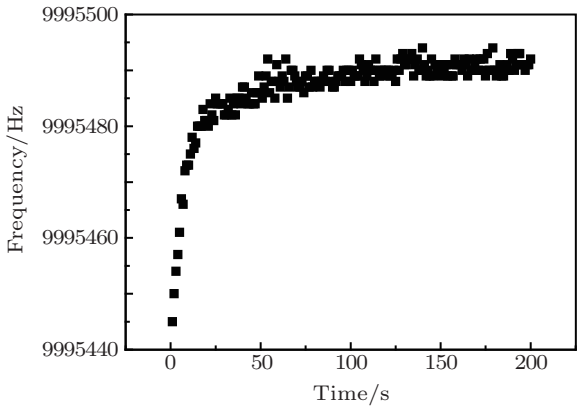


Fig. 1. Signal distribution of the cleaning stage.

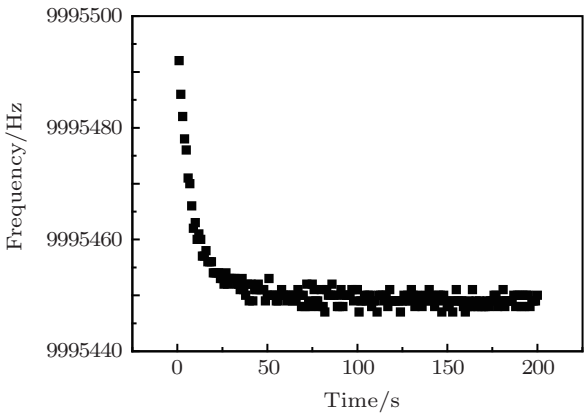


Fig. 2. Signal distribution of the absorption stage.

The ingredients in Chinese spirits are complex, its volatile gas not just a simple overlay of a chemical ingredients. It is impossible to detect volatile gases of Chinese spirits by using the single sensor, therefore, we use eight different kinds of new sensors to compose a sensor array. We could monitor the change in frequency of the sensor array and capture the data of the change in frequency by using a frequency counter. The sensor array is the most important component of the new e-nose.

The new e-nose can be used to detect Chinese spirits. We select 11 different kinds of Chinese spirits as shown in Table 1.

Table 1. The 11 kinds of Chinese spirits.

No	Name
1	Shansi spirits
2	GuoJiao1573
3	Safflower Lang
4	Jiannanchun
5	Jiugui Liquor
6	Kouzi cellar
7	Moutai
8	Niu Lanshan
9	Shuijiangfang
10	Wuliangye (spirit)
11	Xifeng Liquor

Through the chemical composition analyses of the 11 kinds of Chinese spirits, eight kinds of sensors made from eight kinds of nonmetallic polymer thin film coatings shown in Table 2 are obtained. The 8-channel sensing unit of the new e-nose is obtained.

Table 2. Coating materials on different sensors of the sensor array.

Sensor number	Coating materials
1	PVC
2	polyamide
3	polyethylene+AgCl
4	polytef
5	AgCl
6	azithromycin
7	$CuCl_2$+PE
8	$CuCl_2$+AgCl+PE

Using the 8-channel new e-nose, we carry out 50 experiments for each of the spirits, and obtain 550 groups of original signals. If the number of the original signals is large enough, the template data will accurately indicate the characteristic of the test sample.

The working process of the new e-nose for detecting characteristic material contains two phases: the cleaning phase and the adsorption phase. We are concerned with the change rule in the frequency of the new sensor in the adsorption phase. The change rule in the frequency reflects the characteristic information about the Chinese spirits. A new pattern recognition system can process the characteristic information, which we designed.

3. New pattern recognition system

The new pattern recognition system we designed is the most important part of the new e-nose. There are four basic principles for designing the new pattern recognition system. First off, we select a stable interval from the original signal and filter the stable interval using the moving average, thereby obtaining the characteristic value of the original signal. Secondly, the characteristic value will be mapped into the identified space using LDA, whereby we can obtain the feature points that will make up the template data for the new e-nose. Thirdly, the new e-nose will be employed to process the original signal of the test sample through using the new pattern recognition system, and thus, we can acquire the feature point of the test sample. Fourthly, using Euclidean distance to compare the feature point of the test sample with the template data, we can obtain the detection results.

In the signal preliminary processing stage, we select a stable interval from the original signal with the help of Origin software and filter the stable interval using the moving average, thereby obtaining the characteristic value of the original signal.

As shown in Fig. 3, according to the original signal's tendency of distribution, we know that as the absorption time is prolonged, the signal value of the sensor decreases at first, then levels out. We use Origin software to fit the frequency value, the fitting figure shown in Fig. 3, while the fitting curve as a piecewise function is shown in Eq. (6).

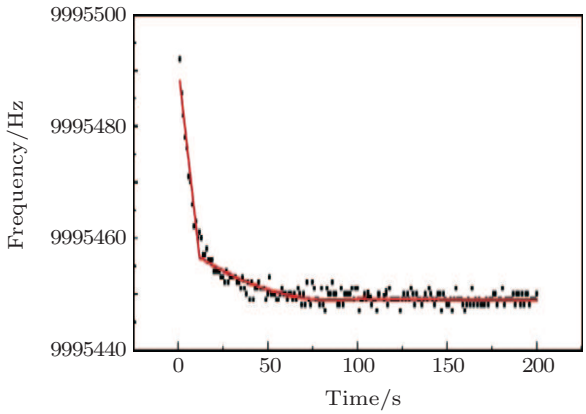


Fig. 3. (color online) Fitting figure of frequency versus time.

$$\begin{cases} y = -2.8656x + 9995490.824, \\ \quad (0 < x \le 12\ \text{s}), \\ y = 0.00136x^2 - 0.23658x + 9995459.08, \\ \quad (12\ \text{s} \le x \le 100\ \text{s}), \\ y = 9995449 \quad (x \ge 100\ \text{s}), \end{cases} \tag{6}$$

where x represents time in seconds, and y represents frequency in Hz. For the piecewise function, the changes in frequency

value display a linear trend in a range from 0 (s) to 12 (s). In a range from 12 (s) to 100 (s), frequency values display a nonlinear trend as time passes by. After 100 (s), frequency values reach a steady state. The steady stage demonstrates that the sensor has reached a condition of total saturation. Therefore, we could obtain a stable interval from 100 (s) to 200 (s).

Based on this principle, we can obtain the other 549 series stable intervals in the absorption phase. From 100 (s) to 200 (s), 100 frequency values can be obtained. These values are the source data of the new pattern recognition system. We divide the 100 frequency values into 10 sequential units. Each unit contains 10 source data. Given a series of numbers and a fixed subset size, the first element of the moving average is obtained by taking the average of the initial fixed subset of the number series. Then the subset is modified by "shifting forward"; that is, excluding the first number of the series and including the next number following the original subset in the series. This creates a new subset of numbers, which is averaged. This process is repeated over the entire data series, after which, we can obtain the eigenvalue of the stable interval.

Each Chinese spirit has eight dimensions of eigenvalue and 550 groups of original signals have eight dimensions and 11 kinds of classification.

The algorithm of the moving average can negate the uncertain element. With the new pattern recognition system, the combining multi-order matrix LDA is crucial for detection results. The new pattern recognition system can reorganize the feature-extracting results using a multi-order matrix LDA algorithm that can convert the 8-dimensional eigenvalue into low dimensions.

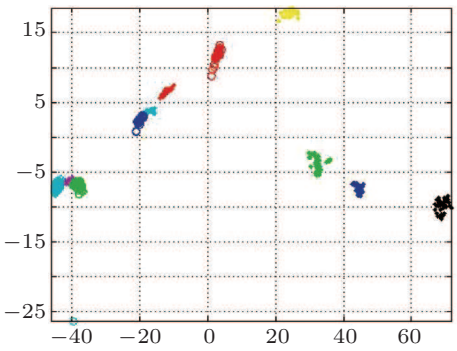


Fig. 4. (color online) Template data of 2-dimensional projection space.

This process can be recorded with Matlab software. We simulate the new pattern recognition system, process the 8-dimensional eigenvalue, and set up the template data for the 11 kinds of Chinese spirits. In order to distinguish between the 11 kinds of Chinese spirits, we convert the 8-dimensional eigenvalue into 2- or 3-dimensional ones to realize the projection from the eigenvalue to the recognition space. The projection results are shown in Figs. 4 and 5. In the images, each point represents the projection point of the 8-dimensional eigenvalue for one experiment, while the different colors designate the different kinds of Chinese spirits. From the figures, we can surmise that different Chinese spirits will fall into their corresponding areas.

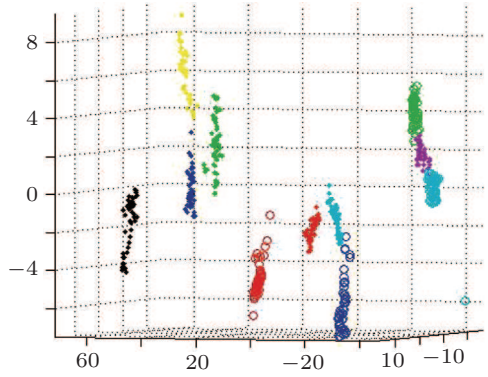


Fig. 5. (color online) Template data of 3-dimensional projection space.

As the images show, the template data of 11 kinds of Chinese spirits cannot be distinguished in the 2-dimensional projection space. However, in the 3-dimensional projection space, we can easily distinguish between the different kinds of Chinese spirits.

In order to find a reasonable dimension that will provide a better result of separation, we need to confirm the dimension of the projection space according to the result of separation. Due to the fact that the 3-dimensional projection results can distinguish the 11 kinds of Chinese spirits, we select the data of the 3-dimensional projection space to use as the template data, and they form a scatter diagram of 50×11 points. We process the 50 template data by the method of weighted average of 1, and gain 11 kinds of Chinese spirit template data points (x_i, y_i, z_i) $i = 1,1,2,\ldots,11$.

The back-end circuit of the new e-nose could capture the original signal of any of the 11 kinds of Chinese spirits above. With the processing of the new pattern recognition system to the original signal, we obtain the sample data (x, y, z) of this spirit. The results of the experiments are based on the following steps. First, the eigenvalue of the sample data is compared with the template data; second, the Euclidean distances between the sample data and the 11 kinds of template datum are found; third, we process the Euclidean distance by the normalized index of transformation; and finally, we sort the results according to the values and obtain the optimal similarity.

The Euclidean distance between the sample source data $(x,\ y,\ z)$ and 11 kinds of template data $(x_i,\ y_i,\ z_i)$ $i = 1,1,2,\ldots,11$ are

$$d_i = (x-x_i)^2 + (y-y_i)^2 + (z-z_i)^2, \quad i = 1,1,2,\ldots,11. \quad (7)$$

Disposing the Euclidean distance by the normalized index

of transformation, we gain the value by similarity

$$P = \mathrm{EXP}(-d_i/N) \times 100\%; \quad \text{where} \quad N = 400. \tag{8}$$

Sorting the test results based on the value P, the value N is obtained by the Euclidean distances of 11 kinds of liquors. Due to the fact that the largest Euclidean distance value of all kinds of liquor is 400, we set the value $N = 400$.

4. Experiments

In order to determine whether the new pattern recognition system is effective, we carry out a large number of experiments monitoring it. GouJiao1573, HongHualang, and Moutai are selected as the sample spirits, the experiments are carried out using the new e-nose at room temperature.

After gathering the original signals of the samples using the new e-nose, we process them via the new pattern recognition system and acquire the feature point for each test in the 3-dimensional space. Then the Euclidean distances between the feature point with 11 kinds of template datum can be obtained. We process the Euclidean distances by normalized index of transformation and obtain 11 percentages. The percentages are called similarity. Finally, we sort the results according to the percentages and obtain the optimal similarity.

Four radar charts of similarity are shown in Fig. 6. In the radar chart, the eight lines represent the response values for the 8-channel sensing unit. In the figure, we select four radar charts that have the largest similarity among the 11 kinds of Chinese spirits. The response values of each test sample are then compared with those of the four kinds of Chinese spirits.

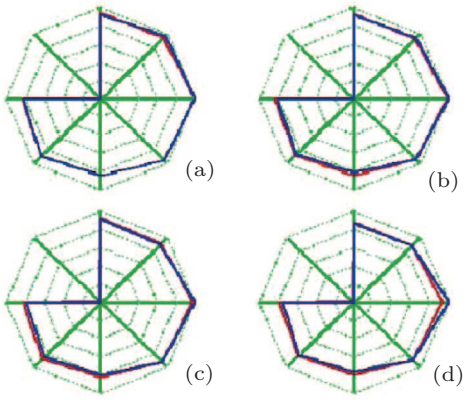


Fig. 6. (color online) Radar charts of optimal similarity.

The results of the three samples can be seen in Table 3.

Table 3. Detecting results.

No	Case 1 Name of liquor	Case 1 Similty P (100%)	No	Case 2 Name of liquor	Case 2 Similty P (100%)	No	Case 3 Name of liquor	Case 3 Similty P (100%)
1	HongHualang	98.5	1	GouJiao1573	99.2	1	Moutai	99.4
2	GouJiao1573	77.4	2	HongHualang	89.2	2	HongHualang	52.3
3	Moutai	59.4	3	Kouzi cellar	61.5	3	Shansi spirits	24.3
4	Kouzi cellar	40.0	4	Niu Lanshan	26.5	4	Kouzi cellar	9.6

As the above results exhibit, these three kinds of Chinese spirits are distinctly different from the others, and the accuracy rate of similarity is higher than 98%. These results convince us that the new pattern recognition system can process the original signal of the 8-channels new e-nose when detecting 11 kinds of Chinese spirits, and it can also distinguish between the different Chinese spirits.

5. Conclusion

In this paper, we show that the new pattern recognition system based on the moving average and LDA can be used to process the original signals of the new polymer quartz piezoelectric crystal air-sensitive sensor system we designed, which is called the new e-nose. With the new pattern recognized, the new e-nose can process the original signals of 11 kinds of Chinese spirits and obtain the template data for detecting the type of Chinese spirits. The 3 kinds of Chinese spirits selected can validate the effectiveness of this new pattern recognition system, and the detected results illustrate that the new pattern recognition system can perfectly distinguish between the 11 kinds of Chinese spirits.

References

[1] Gu Y, Li Q, Tian F F and Xu B J 2013 *Chin. Phys. Lett.* **30** 108101
[2] Muhammad T S C, Karimov K S, Khalid F A, Abbas S Z and Bhatty M B 2013 *Chin. Phys. B* **22** 010701
[3] Sekhar P K, Brosha E L, Mukundan R, Linker K L, Brusseau C and Garzon F H 2011 *J. Hazard Mater.* **190** 125
[4] Ruys D P, Andrade J F and Guimar aes O M 2000 *Analytica Chimica Acta* **404** 95
[5] Ng T N, Wong W S, Chabinyc M L, Sambandan C S and Street R A 2008 *Appl. Phys. Lett.* **92** 213303
[6] Steibig H, Street R A, Knipp D, Krause M and Ho J 2006 *Appl. Phys. Lett.* **88** 013509
[7] Guo D M, Zhang D and Zhang L 2011 *Sensor Actuat B-Chem.* **157** 265
[8] Potyrailo R A and Morris W G 2007 *Rev. Sci. Instrum.* **78** 072214
[9] Kildishev A V, Kamondetdacha R and Nyenhuis J A 2003 *J. Appl. Phys.* **93** 7074
[10] Gao Z K, Zhang X W, Jin N D, Norbert M and Kurths J 2013 *Phys. Rev. E* **88** 032910