International Journal of u- and e- Service, Science and Technology
Vol.8, No. 8 (2015), pp.55-64
http://dx.doi.org/10.14257/ijunesst.2015.8.8.06

The Research on the Multi-Sensor Information Fusion Identifying of Alcohol based on Modified PCA and ANN

Liang Zhao[1], Jun Lu[2], Deng-Feng Chen[1] and Wei Wang[3]

[1]*Dept of Info and Automation, Xi'an Univ. of Arch. &Tech., Xi'an 710055, China*

[2]*School of Automation and Information Engineering, Xi'an University of Technology, Xi'an 710048, China*

[3]*Schoolt of Civil Engineering, Xi'an Univ. of Arch. &Tech., Xi'an 710055, China*

Macusezhao@gmail.com

Abstract

Electronic nose (EN) is a equipment with ability of identification of simple or complex odors. Because of its low cost and accurate identification rate, the researches of it attract more attention and it develops quickly. Firstly, through Gas Sensor Array, we get large amount sample data and all data which is preprocessed. Secondly, modified Bp algorithm and RBF algorithm combining nearest neighbor - clustering algorithm and K-means clustering algorithm is proposed to realize the identification. Principal component analysis (PCA) method wipes off redundant sensor from the sensor array. Principle components and brief sensor signals are tested by above two algorithms. The test result indicated that the rapid and exact identification measure of ANN combining PCA is provided to the pattern identification with sensor information fusion.

Keywords: *PCA, nearest neighbor –K means clustering RBF algorithm, modified BP algorithm, Gas Sensor Array*

1. Introduction

Currently liquor identification has become an artificial olfactory systems research direction. Liquor is a mixture of a variety of chemical components, the alcohol is the main component, further comprising water and a large number of chemical substances. These chemicals can be divided into types of alcohols, aldehydes, acids, esters, and et al. Determine the quality of the liquor ingredients are often very low levels, but the species is very much. Because Liquor is formed from many single gases, so liquor more complex than a single gas. H. V. Shurmer and J. W. Gardner [1-4] used 12 kinds of metal sensors on five kinds of liquor were classified, and achieved better results. In this paper, multi-sensor array and information fusion technology are combined. Improved Bp algorithm and RBF algorithm (nearest neighbor - clustering algorithm and K-means clustering algorithm) combine Principal component analysis (PCA) method to classify different varieties of liquor.

2. Experiment System Model

Since the operating characteristics of a gas sensor having a non-linear and non-single selective, identification a gas or mixture gas concentration with single gas concentration sensor is difficult to classify [5]. So we adopt different types of gas sensor to constitute sensor array. The detecting data of sensor array will be preprocessed and analyzed through pattern recognition processing or neural networks and PCA. Figure 1 is sensor

ISSN: 2005-4246 IJUNESST

array detecting system model. The hardware system is mainly comprised five parts: test vessel, the sensor array circuit board, A/D converter, computing and power.

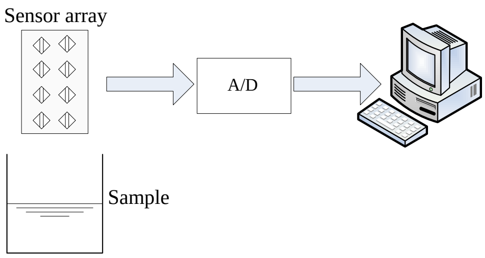


Figure 1. Sensor Array Test System Model

The experimental device uses a sensor array of semiconductor sensors which assemble the front of artificial olfactory system sensing portion. Sensor array contains four kinds of sensors of Figaro company, two kinds of sensors NEMETO company and Honeywell temperature and humidity sensors. The output signal of the sensor array through the analog to digital converter (ADAM-4017) is converted to a digital signal. Because ADAM-4017 using RS485 protocol, so we must through RS485-RS232 converter connected to the computer serial port. The computer can get the sensor array output signal.

3. PCA and Modified ANN Algorithm

3.1. PCA (Principal Component Analysis)

Principal component analysis is the use of dimensionality reduction ideas which adopt multivariate statistical analysis to convert the multi-index to a few comprehensive index [6]. Data matrix is *X*, which conclude specify test data. Principal component PC_k is as follow:

$$PC_k = a_{1k}X_{1j} + a_{2k}X_{2j} + a_{3k}X_{3j} + \cdots + a_{nk}X_{nj} \quad (1)$$

k is principal component label, *n* is the number of dimensions of data. In PCA studying, after the original matrix of data is standardized, analyze the matrix and get principal component of data [7].

3.2. Modified BP and RBF Algorithm

Modified BP algorithm [8] adopts variable step size and adds factor γ^{i}_{pk} .*p* is input vector, *k* is node, *i* is layer. If $\left(t_{pk} - o^{i}_{pk}\right) \neq 0$, the second category local minimum is appeared. t is target value. We hope error function of local or flat regional to change. o^{i}_{pk} rapid exit dead zone, and add a factor γ^{i}_{pk} .Real output of node k is:

$$o^{i}_{pk} = \frac{1}{1+\exp(-\alpha)} = \frac{1}{1+\exp[-\sum_k \omega^{i}_{kj} o^{h}_{pj} / \gamma^{i}_{pk}]} \qquad (2)$$

For each pk and o which enter second category local minimum or flat region, we will make ω^{i}_{kj} reduce a factor and $\gamma^{i}_{pk} > 1$ .So o^{i}_{pk} will leave flat region and $\gamma^{i}_{pk} = 1$. The method can effective avoid second category local minimum and improve convergence speed.

3.3. Nearest K-means Clustering Algorithm

Nearest k-means clustering algorithm which is combined form nearest neighbor clustering algorithm and k-means clustering algorithm is proposed to select RBF basis function. Since the algorithm is adaptive clustering algorithm [9]. So it needn't the number of hidden units which are already confirmed. The conclusion of clustering RBF is optimal. The training regulation is:

(1) Select a appropriate Gauss function width *r*. Define a vector *A*(*l*) which storage various input sum of vector. Define a counter *B*(*l*) which statistics number of various sample. *l* is number of category.

(2)Start from the first data x_1 , establish a clustering center on x_1.If $c_1 = x_1$, $A(l) = c_1$, $B(l) = 1$, only one hidden unit on RBF. The center of hidden unit is c_1.

(3) If data x_k of *k* sample is appear, *k* =2, 3 ,4 ,… , *N*, so exist *M* clustering center which center is $c_1, c_2, \ldots, c_M$ .This also means existing *M* hidden unit. Then respectively calculate distance $| x_k - c_i |$ from x_k to clustering center of *M*. *i* =1 ,2 , … , *M*. Suppose $| x_k - c_j |$ is minimum distance, which c_j is nearest neighbor clustering with x_k .

So, if $| x_k - c_j | > r$, x_k is new center of clustering. Define $c_{M+1} = x_k$, $M = M + 1$, $A(M) = x_k$, $B(M) = 1$ 。 And keep data of *A*(*i*),*B* (*i*).Add *M* hidden units in established RBF. If $| x_k - c_j | \le r$, then calculate $A(j) = A(j) + x_k$, $B(j) = B(j) + 1$, $c_j = A(j) / B(j)$ 。 If *i* ≠*j*, *i* = 1 ,2 , …, *M* , and keep data of *A*(*i*),*B* (*i*).

(4)After select center according regular, calculate normalization parameter ${\sigma_j}^2$ which indicates relation of data on each center. It equal average distance between clustering center and training pattern.

$$ {\sigma_j}^2 = \frac{1}{M_j} \sum_{x \in \theta_j} (x - c_j)^T (x - c_j) \qquad (3) $$

M_j is θ_j pattern number.

Hidden unit output:

$$ G_j(x) = \exp[-\frac{\| x - c_j \|}{2{\sigma^2}_j}] \qquad (4) $$

The algorithm is composed of nearest neighbor clustering algorithm and k- means clustering algorithm. Radius r determines the size of the complexity of the dynamic adaptive RBF network. *r* is smaller, the number of clusters obtained by the more, the greater the amount of computation. But because *r* is a one-dimensional parameter, you can usually find a suitable *r* through experiments and error information, which is more convenient than the same time to determine the number of hidden units and an appropriate norm [10]. Since each input - output data may generate a new cluster. Therefore, this dynamic adaptive RBF network, in fact, parameters and structure of the two processes are adapted in the same time.

4. Experiment and Result Analysis

Experiment equipment collect two kinds of Chinese liquor (Hongxing 500 and Taibai) signal to test. In the same temperature ($25^{\circ}C$), acquire steady-state response signal values of gas sensors for air. Sample sampling is 120s, interval 300s. After gas sensors recover steady-state response values for air, the next sample will be collected.

Each liquor samples of 20 groups. 20 kinds of samples with different concentrations can be taken from 5% concentration liquor which separated from 100% concentration liquor. Two kinds of liquor have 40 groups' samples. Take the average of 12 samples of each of the liquor sensor signal as the training samples, the average of 8 samples of the sensor signal as a measurement of the test sample. Training sample set is 24; the test sample set is 16.

4.1. Data Preprocessing and Data Sample

In order to improve the classification performance of the system identification, the data samples are preprocessed to eliminate or reduce the output effect of the gas concentration sensor.

$$x_{gas,i} = \frac{G_{gas,i}/G_{air,i}}{\left[\frac{1}{n}\sum_{i=1}^{n}\left(G_{gas,i}/G_{air,i}\right)^2\right]^{\frac{1}{2}}} \tag{5}$$

G is hidden output. i from 1 to 6 and x is voltage of gas sensor.

Table 1. Data of Gas Sensors Array

Sample	$x_{gas,1}$	$x_{gas,2}$	$x_{gas,3}$	$x_{gas,4}$	$x_{gas,5}$	$x_{gas,6}$	liquars
1	0.2942	0.2939	0.2932	0.296	0.2991	-2.3588	Taibai
2	0.2707	0.2702	0.2699	0.2735	0.2761	-2.3727	Taibai
3	0.2727	0.2722	0.2720	0.2750	0.2785	-2.3716	Taibai
4	0.2778	0.2773	0.2766	0.2798	0.2810	-2.369	Taibai
5	0.2779	0.2765	0.2760	0.2792	0.2817	-2.3691	Taibai
6	0.3117	0.3121	0.3113	0.3147	0.3166	-2.3472	Taibai
7	0.3358	0.3349	0.3339	0.3373	0.3400	-2.3311	Taibai
8	0.3467	0.3458	0.3449	0.3490	0.3487	-2.3233	Taibai
9	0.3711	0.3702	0.3692	0.3733	0.3760	-2.3040	Taibai
10	0.3573	0.3565	0.3562	0.3605	0.3625	-2.3145	Taibai
11	0.3548	0.3540	0.3552	0.3592	0.3615	-2.3158	Taibai
12	0.3812	0.3805	0.3797	0.3822	0.3849	-2.2960	Taibai
13	0.3671	0.3666	0.3660	0.3702	0.3727	-2.3067	Taibai
14	0.4292	0.4281	0.4274	0.4319	0.4343	-2.2527	Taibai
15	0.4984	0.4973	0.4964	0.5020	0.5044	-2.1798	Taibai
16	0.5025	0.5037	0.5025	0.5084	0.5106	-2.1730	Taibai
17	0.4900	0.4892	0.4879	0.4935	0.4945	-2.1896	Taibai
18	0.6806	0.6796	0.6787	0.6862	0.6891	-1.9153	Taibai
19	0.5683	0.5699	0.5688	0.5752	0.5776	-2.0891	Taibai
20	0.6658	0.6650	0.6638	0.6708	0.6719	-1.9423	Taibai
21	0.8189	0.8154	0.8079	0.8021	0.8096	-1.6471	Hongxing
22	0.4862	0.4843	0.4829	0.4850	0.4884	-2.1959	Hongxing
23	1.0057	0.9984	0.9923	0.9857	0.9954	-1.0221	Hongxing
24	1.0005	0.9927	0.9854	0.9801	0.9876	-1.0519	Hongxing
25	1.0049	0.9978	0.9919	0.9871	0.9967	-1.0213	Hongxing
26	0.5970	0.5946	0.5904	0.5932	0.5969	-2.0575	Hongxing
27	0.9137	0.9072	0.9020	0.9002	0.9080	-1.3761	Hongxing
28	0.9140	0.9081	0.9032	0.9002	0.9073	-1.3750	Hongxing
29	0.5741	0.5723	0.5699	0.5724	0.5785	-2.0871	Hongxing
30	0.9346	0.9285	0.9237	0.9208	0.9305	-1.3029	Hongxing
31	0.6632	0.6604	0.6580	0.6607	0.6685	-1.9513	Hongxing
32	0.7691	0.7663	0.7629	0.7659	0.7685	-1.7499	Hongxing
33	0.8060	0.8053	0.8016	0.8052	0.8085	-1.6605	Hongxing
34	0.8441	0.8413	0.8385	0.8425	0.8470	-1.5651	Hongxing
35	0.8313	0.8288	0.8258	0.8291	0.8341	-1.5991	Hongxing
36	0.7012	0.6989	0.6966	0.6998	0.7032	-1.8842	Hongxing

37	0.8179	0.8150	0.8118	0.8151	0.8185	-1.6351	Hongxing
38	0.9434	0.9401	0.9365	0.9377	0.9424	-1.2577	Hongxing
39	1.0695	1.0655	1.0615	1.0659	1.0648	-0.5693	Hongxing
40	1.0807	1.0772	1.0725	1.0768	1.0783	-0.4462	Hongxing

4.2. PCA Research

After PCA (Principal component analysis) of normalized samples**,** the result is as show Figure 2.

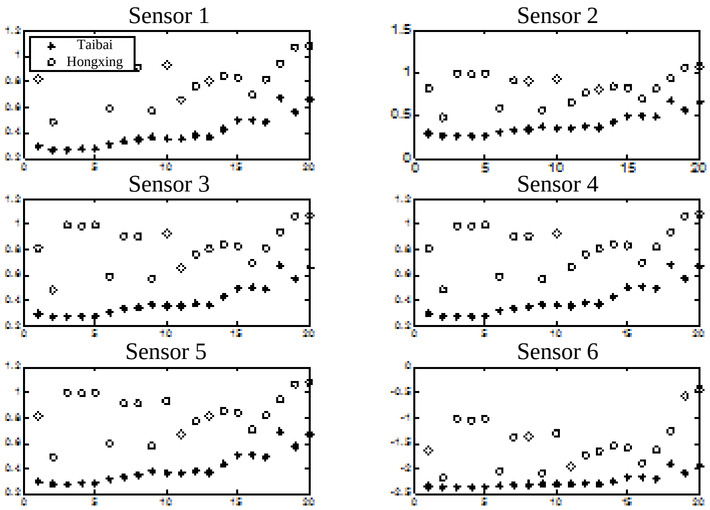


Figure 2(a). Outputs of Six Sensors

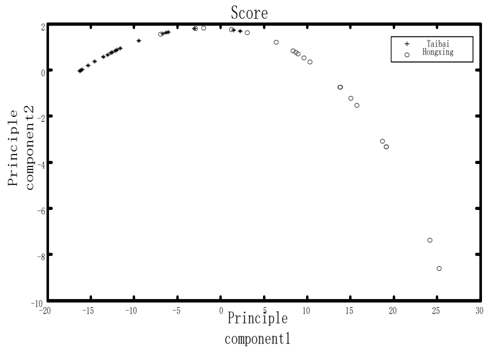


Figure 2. (b) Score Plot of 3 Samples in the PC1-PC2

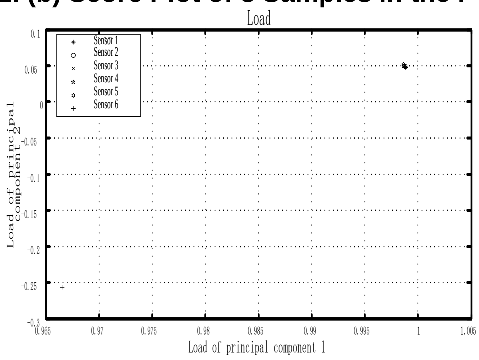


Figure 2. (c) Load Plot of 3 Samples in the PC1-PC2

Figure 2. Principle Component Analysis

From Figure 2(b), 2 kinds of liquor exit overlap and not good classification. In load Figure, similar load represents redundancy. There exit similar load which means exiting redundancy information in original data of sensor array. So we can remove some overlap data from sensor 1 to sensor 5 of sensor array. This form experiment data from 2 sensors (sensor 1 and sensor 6). But the effect of classification of PCA is not good, so we will adopt modified ANN to analyze signal. The input data which consist of the original data of 6 sensors signal, removed redundancy information of 2 sensors signal and 2 principle component data is classified.

4.3. Modified ANN Research

After 206217 times training of modified BP, 2 samples is error classified. Error converges is 0.01.The result of modified BP training is as show Figure 3 and Table 2.

Figure 3. Error and Training Step of Modified BP Algorithm

Table 2. The Identification Result of BP Algorithm

Sample serial	output	recognition results	actual liquor
2	0.0000 1.0000	Taibai	Taibai
4	0.0000 1.0000	Taibai	Taibai
6	0.0000 1.0000	Taibai	Taibai
8	0.0000 1.0000	Taibai	Taibai
10	0.0000 1.0000	Taibai	Taibai
12	0.0000 1.0000	Taibai	Taibai
14	0.0000 1.0000	Taibai	Taibai
16	0.0000 1.0000	Taibai	Taibai
22	0.0000 1.0000	Hongxing	Taibai
24	1.0000 0.0000	Hongxing	Hongxing
26	1.0000 0.0000	Hongxing	Hongxing
28	1.0000 0.0000	Hongxing	Hongxing
30	1.0000 0.0000	Hongxing	Hongxing
32	0.9961 0.0038	Hongxing	Hongxing
34	1.0000 0.0000	Hongxing	Hongxing
36	0.0198 0.9775	Hongxing	Taibai

The classification result of nearest neighbor K-means clustering algorithm of modified RBF is same Table 2. But training times is 31943 and error is 0.7. The result of nearest neighbor K-means clustering algorithm of modified RBF training is as show Figure 4.

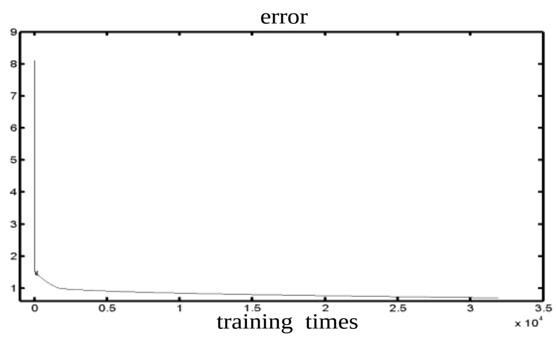


Figure 4. Error and Training Step of Modified RBF Algorithm

4.4. Result Reduce the Data Dimension of PCA

Modified BP algorithm training sample of sensor 5 and 6.After through 12869 times training, the error is 0.001.After through 12196 times training of principle component 1 and component 2,the error is 0.001.Although output of classification on net is different, result is same on two ways. The Table 3 is as show result of modified BP network with 2 PCs data as input. Figure 5 is as show err and training times of modified BP algorithm.

Table 3. The Identification Result of Modified BP Network with 2 PCs Data as Input

Sample serial	output	recognition results	actual liquor
2	0.0002 0.9999	Taibai	Taibai
4	0.0002 0.9999	Taibai	Taibai
6	0.0002 0.9999	Taibai	Taibai
8	0.0002 0.9999	Taibai	Taibai
10	0.0003 0.9998	Taibai	Taibai
12	0.0003 0.9998	Taibai	Taibai
14	0.0006 0.9996	Taibai	Taibai
16	0.0030 0.9981	Taibai	Taibai
22	0.0018 0.9989	Taibai	Taibai
24	0.9999 0.0001	Hongxing	Hongxing
26	0.0508 0.9648	Hongxing	Taibai
28	0.9999 0.0001	Hongxing	Hongxing
30	0.9999 0.0001	Hongxing	Hongxing
32	0.9940 0.0072	Hongxing	Hongxing
34	0.9996 0.0005	Hongxing	Hongxing
36	0.8872 0.1415	Hongxing	Taibai

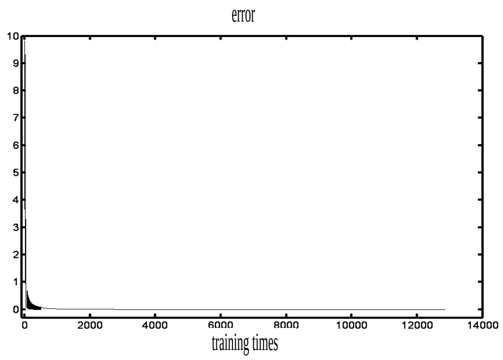


Figure 5(a). 2 Sensors Data as Input

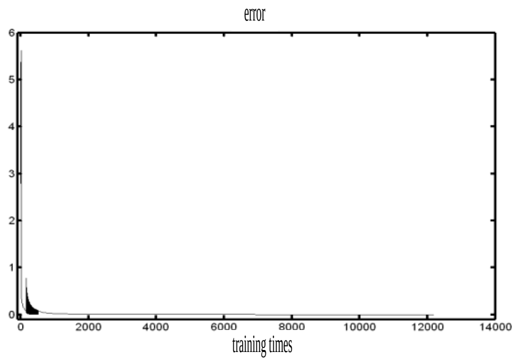


Figure 5(b). 2 PCs Data as Input

Figure 5. Error and Training of Modified BP Algorithm

Nearest neighbor K-means clustering algorithm of modified RBF algorithm training sample of sensor 5 and 6.After through 136869 times training, the error is 0.07.After through 58758 times training of principle component 1 and component 2,the error is 0.07.Although output of classification on net is different, result is same on two ways. The Table 4 is as show the identification result of RBF network with 2 kinds of input samples. Figure 6 is as show err and training times of modified RBF algorithm.

Table 4. Identification Result of RBF Network with 2 Kinds of Input Samples

Sample serial	actual liquor	Output of 2 PCs data as input	recognition results	Output of 2 sensors data as input	recognition results
2	Taibai	0.0275 0.9718	Taibai	0.9398 0.0735	Taibai

4	Taibai	0.0014 0.9985	Taibai	0.2343 0.7980	Taibai
6	Taibai	0.0000 1.0000	Taibai	0.0001 0.9999	Taibai
8	Taibai	0.0000 1.0000	Taibai	0.0000 1.0000	Taibai
10	Taibai	0.0000 1.0000	Taibai	0.0000 1.0000	Taibai
12	Taibai	0.0000 1.0000	Taibai	0.0000 1.0000	Taibai
14	Taibai	0.0000 1.0000	Taibai	0.0000 1.0000	Taibai
16	Hongxing	0.9753 0.0187	Taibai	0.9929 0.0055	Taibai
22	Taibai	0.0358 0.9575	Hongxing	0.0002 0.9998	Taibai
24	Hongxing	1.0000 0.0000	Hongxing	1.0000 0.0000	Hongxing
26	Taibai	1.0000 0.0000	Hongxing	1.0000 0.0000	Hongxing
28	Hongxing	1.0000 0.0000	Hongxing	1.0000 0.0000	Hongxing
30	Hongxing	1.0000 0.0000	Hongxing	1.0000 0.0000	Hongxing
32	Hongxing	0.8985 0.0602	Hongxing	1.0000 0.0000	Hongxing
34	Hongxing	1.0000 0.0000	Hongxing	1.0000 0.0000	Hongxing
36	Hongxing	0.9693 0.0129	Hongxing	1.0000 0.0000	Hongxing

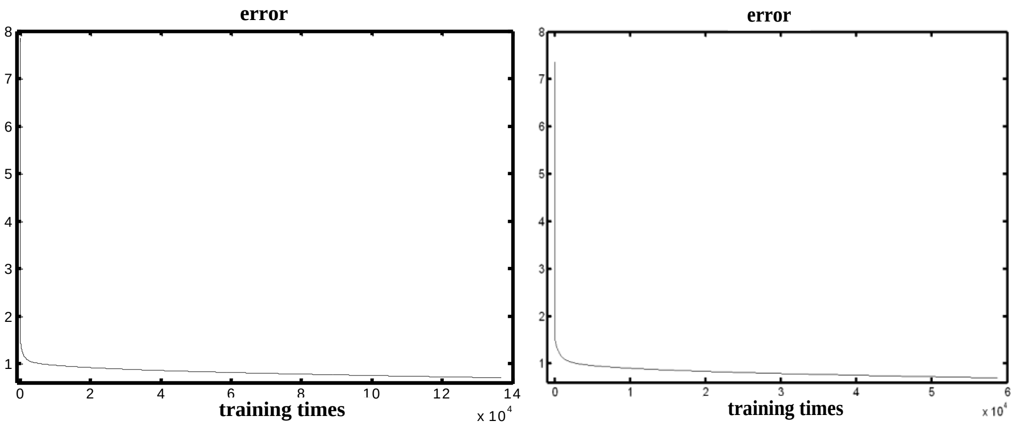


Figure 6(a). 2 Sensors Data as Input **Figure 6(b). 2 PCs Data as Input**

Figure 6. Error and Training of Modified RBF Algorithm with 2 Kinds of Input Samples

Compared with the identification results of before and after reduce the dimension of the principal component, we can conclude result and resolution is almost same. So PCA combine modified ANN can be applied to liquor recognition.

5. Conclusions

Through Gas Sensor Array, we get large amount liqour data which is preprocessed. 8 sensors formed 8 dimensional variable original data. Based on original data of 8 dimensional variable, through PCA analysis, we can decreases to 2 dimensions variable. The 2 dimensional variables as input, correct classification rate of sample can achieve 95.83% on modified BP algorithm, the correct classification rate of sample can achieve 83.33% on nearest neighbor K-means clustering modified RBF algorithm. The result of 2 kinds of liquor classification is show PCA combine modified ANN can be well applied to liquor recognition.

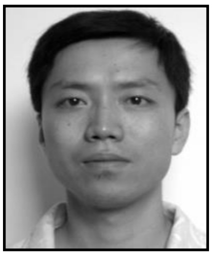
Acknowledgements

This work was financially supported by National Nature Science Foundation of China (Grant No. 51478383, No.5120916), Nature Science Foundation of Shaanxi Province (Grant No. 2013JQ7024), ShaanXi Educational Committee (Grant No.12JK0520).

References

[1] H. V. Shurmer and J. W. Gardner, "Intelligent vapour discrimination using a composite 12-element sensor array", Sensors and Actuators, vol. B1, **(1990)**, pp. 256-260.
[2] H. V. Shurmer and J. W. Gardner, "The application of discrimination techniques to alcohols and tobaccos using tin oxide sensor", Sensors and Actuators, vol. 18, **(1989)**, pp. 359-369.
[3] J. W. Gardner, E. L. Hines and H. C. Tang, "Detection of vapours and odours from a multisensor array using pattern recognition", Sensors and Actuators B, vol. B4, **(1991)**, pp. 109-115.
[4] H. V. Shurmer and J. W. Gardner, "Odours discrimination with an electronic nose", Sensors and Actuators, vol. B8, **(1992)**, pp. 1-11.
[5] M. L. Rodríguez-Méndez, A. A. Arrieta, V. Parra, *et al.*, "Fusion of Three Sensory Modalities for the Multimodal Characterization of Red Wines", IEEE Sensors Journal, vol. 4, no. 3, **(2004)**, pp. 348~354.
[6] M. Pardo and G. Sberveglieri, "Remarks on the Use of Multilayer Perceptrons for the Analysis of Chemical Sensor Array Data", IEEE Sensors Journal, vol. 4, no. 3, **(2004)**, pp. 355~363.
[7] M. Penza and G. Cassaon, "Application of principal component analysis and artificial neural networks to recognize the individual VOCs of methanol/2-propanol in a binary mixture by SAW multi-sensor array", Sensors and Actuators B, vol. 89, **(2003)**, pp. 269-284.
[8] S. A. Vasanwala, J. A. Desai and H. S. Patil, "BP and RBF neural networks for predicting minimum weight of double-layer space grids", International Journal of Applied Engineering Research; vol. 3, no. 7, **(2008)**, pp. 969-979.
[9] R. Savitha, S. Suresh and N. Sundararajan, "fully complex-valued radial basis function network and its learning algorithm", Int. J. Neur. Syst. vol. 19, no. 14, **(2009)**, pp. 253-.265.
[10] N. Lopes and B. Ribeiro, "An evaluation of multiple feed-forward networks on GPUs", Int. J. Neur. Syst., vol. 21, no. 1, **(2011)**, pp. 31-45.

Authors



Liang Zhao, was born 1980. He completed his Ph.D. degree in Xi'an University of Architecture and Technology in China. He research focus on computer control, measure and detection technique.

Jun Lu, received M.S degree from XiDian University in China. Her research interests in computer control, information system.

Deng-feng Chen, completed his Ph.D. degree in Northwestern Polytechnical University in China. He mainly focus research on computer detection technique, control engineering.

Wei Wang, received Ph.D. degrees in Xi'an University of Architecture and Technology in China. He is currently a associate professor in School of Civil Engineering at Xi'an University of Architecture and Technology in China. He mainly focus research include Building Structure Design and Safety Monitor & Diagnosis.

