DOI 10.5162/IMCS2012/P1.9.22

An Electronic Nose Recognition Algorithm Based on PCA-ICA Preprocessing and Fuzzy Neural Network

Weiping Yan[1], *Zhenan Tang*[1], *Guangfen Wei*[2], *Jun Yang*[1]
[1] *School of Electronic Science and Technology, Dalian University of Technology, 116024,Dalian, Liaoning Province, P.R. China*
[2] *School of Information and Electronics, Shandong Institute of Business and Technology, 264005, Yantai, Shandong Province, P.R. China*
e-mail: yanwp@dlut.edu.cn

Abstract:

To improve the recognition performance of electronic noses detecting gas mixtures, a PCA-ICA signal preprocessing and fuzzy neural network based recognition algorithm is proposed. In this approach, signals of electronic noses are firstly preprocessed effectively by combination of Principal Component Analysis (PCA) and Independent Component Analysis (ICA), and then processed with a fuzzy Takagi-Sugeno system integrated with multi neural networks for the purpose of quantification of gas concentrations. Experiment results show that the alcohol concentration recognition performance is highly improved in alcohol and gasoline mixtures even interfered by smokes.

Key words: gas sensor, electronic nose, fuzzy neural network, Principal Component Analysis, Independent Component Analysis

Introduction

Caused by the physical shortcomings of gas sensors, it is impossible for a single gas sensor to identify multiple gases. So the electronic nose techniques based on gas sensor array and pattern recognition is becoming an important way of dealing with cross-sensitivity in gas analysis. The pattern recognition technology plays a crucial role on the behavior of electronic noses. Aimed to this requirement, most pattern recognition techniques have been widely studied to realize the qualitatively or quantitatively recognition of gas mixtures, such as statistical techniques, Artificial Neural Networks and fuzzy systems [1-3]. This paper proposes a PCA-ICA signal preprocessing and fuzzy neural network based electronic nose algorithm, to quantitatively recognize the alcohol concentrations under interferences of gasoline and smokes.

PCA-ICA Based Signal Preprocessing Method

Suppose *N* sensors $(x_1, x_2, \cdots x_N)$ are chosen to construct an array to detect the gas mixtures containing *M* kinds of gases $(s_1, s_2, \cdots s_M)$. If the gas sensor responses are linear, they can be denoted as

$$\begin{cases} x_1 = a_{11}s_1 + a_{12}s_2 + \cdots + a_{1M}s_M \\ \vdots \\ x_N = a_{N1}s_1 + a_{N2}s_2 + \cdots + a_{NM}s_M \end{cases} \tag{1}$$

Or a matrix form

$$\boldsymbol{X} = \boldsymbol{AS} \tag{2}$$

Where X is the N-dimensional vector of sensors, A is an $N \times M$ matrix denoting the cross sensitivity of gas sensor array, S is the M-dimensional vector of gases. This electronic nose model is actually the ICA model without noise. Therefore, ideally the gas information could be obtained based on ICA principle, at least mutual information could be reduced.

However, the number of sensors (*M*) is generally supposed equal to the number of sources (*N*) in ICA. While in electronic nose system, in order to reduce the cross sensitivity between sensors, the number of sensors is always designed larger than the number of sources. Therefore, PCA is adopted firstly to reduce the dimensions. PCA is a mathematical procedure that uses an orthogonal transformation to convert a set of observations of possibly correlated variables into a set of values of linearly uncorrelated variables called principal components. It has been widely used in electronic nose systems. Generally, the obtained gas mixture response data are firstly performed with PCA to reduce the dimensions [4].

Procedures of traditional algorithms are projecting data vectors to the PCA eigen space firstly and processing the mapped new data with ICA secondly [5], which needs repeated projections of high dimensional data. While in this paper, ICA is performed directly to the PCA eigen matrix ***Pm*** to obtain the ICA eigen space ***U***, and the original data is directly projected to the final ICA eigen space. Repeated operations of high dimensional data are avoided and the computation amount is reduced.

The preprocessing algorithm is designed as follows:

(a) Compute the $E\left\{\boldsymbol{X}\boldsymbol{X}^T\right\}=\boldsymbol{C}=\frac{1}{l}\sum_{j=1}^{l}\overline{x_j x_j^T}$.

(b) Diagonalize the matrix, and obtain the eigenvalues and eigenvectors of the correlation matrix E{ $\boldsymbol{X}\boldsymbol{X}^T$ } of input vector ***x*** . Order the eigenvalues, $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$. And the eigenvectors construct an orthorgonal matrix $\boldsymbol{P}=[p_1, p_2, \cdots, p_n]$.

(c) Eigenvectors corresponding to the first ***M*** maximum eigenvalues are chosen out to construct a new matrix Pm.

$$\left(\sum_{i=1}^{m}\lambda_i\right)\Bigg/\left(\sum_{i=1}^{n}\lambda_i\right)\geq 0.85, m<n \qquad (3)$$

(a) After centralization and whitening, the whitened electronic nose PCA eigen matrix ***Pm*** is obtained.

(b) ICA is then applied to the eigen matrix ***Pm*** and the ICA eigen space ***U*** is constructed. $\boldsymbol{U}=\boldsymbol{W}\boldsymbol{P}_{\boldsymbol{m}}$, where W is obtained through the Informax based ICA algorithm.

(c) Finally, the original sample data are mapped to the new eigen space ***U*** directly. $\boldsymbol{Y}=\boldsymbol{U}^T\boldsymbol{X}$, Y is the preprocessed results of original signal $\boldsymbol{X}$.

Neural Network Based Takagi-Sugeno Fuzzy System

Because fuzzy rules mostly are complex nonlinear functions, selection and construction of membership functions are generally subjective as well as the determination of fuzzy inference rules. The neural network-based fuzzy logic system, combined with neural network and fuzzy logic, has the advantages of self-organizing, sample-learning, auto-structuring and fuzzy rule developing and member function optimizing [6]. Therefore, a Takagi-Sugeno fuzzy system based on multi neural network integration is proposed in this paper which is shown in Fig.1.

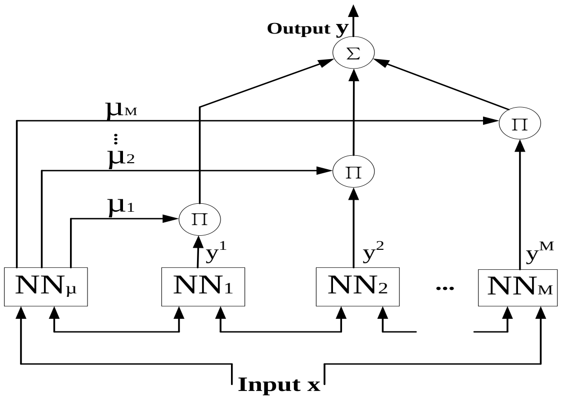

Fig. 1. Takagi-Sugeno fuzzy system based on multi neural networks

There are ***M+1*** neural networks in the recognition system. ***M*** pieces of fuzzy rules are generated through ***NN₁~NN_M*** neural networks, and the applicability of each rule to sample X is generated through the membership association neural network ***NNμ***. The output of Takagi-Sugeno fuzzy system is determined by both the rule output value $\left(y^j\right)$ and the rule membership function $\left(\mu^j\right)$.

The algorithm is designed as follows:

(a) Choose the training sample set.

(b) Pre-classify the training set by Self-Organizing Map (SOM) neural network. Suppose the sample set can be classified to M classes, there will be M fuzzy rules in the fuzzy system. This paper is aimed to recognize the alcohol concentration. Considering the complexity and recognition precision, 3 fuzzy rules are chosen, which denote the high, middle and low concentration of alcohol. Therefore, M=3.

(c) Form the member functions. Using the RBF neural network to train the joint member function neural network ***NNμ***. The number of input of NN_μ is decided by the sensor number. The number of outputs is M, which is decided by step (b). The output of ***NNμ***denotes the membership degree of the input samples to the M rules.

(d) Construct the fuzzy rules. According to the pre-classification results of SOM, fuzzy rule neural networks $NN_1 \sim NN_M$ are built to generate the M fuzzy rules. RBF neural networks are also selected. If $X=(x_1,x_i,\cdots x_n)^T$ is F, then $y^j = \mathrm{NN}_j\ (x_1,x_i,\cdots x_n)$ The output of the neural network denotes the reasoning results.

(e) Obtain the final output $y=\sum_{j=1}^{M}\mu^j \times y^j$.

Where $\mu_i^j=\left\{\mu_1^j\mu_2^j\cdots\mu_n^j\right\}$ is the membership function of *j*th class, $y_i^j=\left\{y_1^j y_2^j\cdots y_n^j\right\}$ is the *j*th class fuzzy rules.

Results and Analysis

Experiments are performed using 8 sensors, including a temperature sensor, a humidity sensor, a smoke sensor and 5 SnO_2 gas sensors. Measurement conditions are set as following: the concentration range of alcohol is 0-500ppm, the concentration range of gasoline is 0-270ppm, temperature range is 20-35℃, humidity range is 20-80%RH, and the concentration of smokes is adequate. In our situations, gasoline is the general interference gas, which is always existed. Smoke is also interference gas, sometimes exists and sometimes not. Responses of the sensor array to different concentrations of alcohol are measured under the static environment and in the situations with different concentrations of gasoline, with or without smokes as the interference gas. The concentration of alcohol is set 100ppm, 150ppm, 200ppm, 250ppm, 300ppm, 350ppm, 400ppm, 450ppm, 500ppm. The concentration of gasoline is set 90ppm, 180ppm and 270ppm. Combining with the smokes, totally 54 kinds of samples are detected. Each sample is tested repeatedly four times. 3 times of test are used as training data, the left one times is used as test data.

The obtained data are processed with the proposed algorithm. After PCA, contributions of the first three principal components are 42.7734%, 33.8442%, 22.2531%. Cumulated contribution occupies 98.8707% of the total information. Therefore, these three principal components are chosen out to construct the feature space. Processed with the Infomax based ICA algorithm, the mapping *U* is obtained and transformed the data to *Y*.

The preprocessed data are then input to the following neural network based Takagi-Sugeno fuzzy system according to above algorithms. The number of the output of pre-classification SOM neural network is designed as 3, which means that the system contains 3 fuzzy rules. By using RBF neural network, the joint membership function and three fuzzy rule neural networks are constructed and trained to obtain the membership functions and reasoning outputs. The RBF neural network is a single hidden layer network, Gaussian function is chosen as the base function, and the expanding rate is set to 2.5, 2.0 and 2.0.

The final output of the fuzzy system is the quantification results. Tab.1 and Tab.2 show the result of situations without smoke interference and with smoke interference. From Tab.1, it can be seen that the concentration of alcohol is precisely recognized with quite low quantification errors. The largest relative error is 0.1626%, the average error is 0.0655%. The error is larger when the gasoline concentration is larger. While in Tab.2, the largest relative error is 0.2362%, and the average error is 0.0854%, which are a little larger than situations without smokes.

Tab.1: The identification result of the system in situations without smokes

Sample	Gasoline (ppm)	Alcohol (ppm)	Output (ppm)	Relative error (%)
1	90	100	99.99	0.0017
2	90	150	149.99	0.0011
3	90	200	200.07	0.0372
4	90	250	249.85	0.0599
5	90	300	299.51	0.1626
6	90	350	349.68	0.0890
7	90	400	400.03	0.0083
8	90	450	450.32	0.0728
9	90	500	500.28	0.0578
10	180	100	100.04	0.0495
11	180	150	150.03	0.0219
12	180	200	200.18	0.0940
13	180	250	249.99	0.0021
14	180	300	300.22	0.0766
15	180	350	350.18	0.0536
16	180	400	399.91	0.0203
17	180	450	449.69	0.0672
18	180	500	500.00	0.0018
19	270	100	99.84	0.1541
20	270	150	149.81	0.1208
21	270	200	200.21	0.1055
22	270	250	249.60	0.1562
23	270	300	300.26	0.0890
24	270	350	350.00	0.0015
25	270	400	399.54	0.1128
26	270	450	450.22	0.0504
27	270	500	499.49	0.1010

Tab.2: Identification result of the system in smog situation

Sample	Gasoline(ppm)	Alcohol (ppm)	Output (ppm)	Relative error (%)
1	90	100	99.97	0.0209
2	90	150	150.00	0.0033
3	90	200	199.91	0.0886
4	90	250	250.06	0.0266
5	90	300	300.01	0.0060
6	90	350	350.04	0.0137
7	90	400	399.97	0.0058
8	90	450	449.84	0.0340
9	90	500	500.33	0.0663
10	180	100	99.95	0.0461
11	180	150	150.18	0.1218
12	180	200	199.85	0.0648
13	180	250	249.86	0.0540
14	180	300	300.15	0.0520
15	180	350	350.34	0.0982
16	180	400	399.53	0.1162
17	180	450	450.23	0.0522
18	180	500	499.59	0.0818
19	270	100	100.20	0.2015
20	270	150	150.35	0.2362
21	270	200	199.63	0.1816
22	270	250	249.47	0.2096
23	270	300	300.47	0.1591
24	270	350	349.71	0.0827
25	270	400	400.29	0.0732
26	270	450	449.85	0.0313
27	270	500	500.06	0.0122

For the purpose of comparison, the general used ICA-BP algorithm is applied to process the data too. In this approach, ICA is used as preprocessing method and a Back-Propagation neural network is adopted as the recognition system [7]. The relative quantification errors of alcohol concentrations are computed from the results of both approaches. Tab. 3 shows the results.

Tab. 3: Comparison of average relative errors

	PCA-ICA, Takagi-Sugeno fuzzy system	ICA-BP system
Without smokes	0.0655%	1.6625%
With smokes	0.0854%	2.1934%

It can be seen from Tab.3 that, better quantification performance is obtained by the PCA-ICA preprocessing and Takagi-Sugeno fuzzy system than the general used ICA-BP system. Especially under situations with smokes, the errors are not increased clearly. The results demonstrate the effectiveness of proposed approach. It seems that the computation amount of the proposed approach is larger, however, because the dimension of data is reduced and redundant information between signals are reduced by the PCA-ICA preprocessing approach, the complexity of structure of the fuzzy system is largely reduced, the recognition speed is faster.

Conclusion

An electronic nose recognition algorithm based on PCA-ICA preprocessing and fuzzy neural network is proposed and described in detail. Adopting this algorithm, the alcohol concentration interfered by gasoline and smokes is recognized. Compared with the traditional ICA-BP algorithm, the proposed approach shows better performance.

Acknowledgements

The authors would like to thank the financial support of National Science Foundation of China (No. 61174007).

References

[1] H.Men, X.Li, J.Wang, J.Gao, IEEE International Conference on Control and Automation, 2699-2704 (2007); doi: 10.1109/ICCA.2007.4376852

[2] J.Brezmes, L.Ma, E.Llobet, X.Vilanova, IEEE Sensors Journal 5(1) 97-108(2005); doi: 10.1109/JSEN.2004.837495

[3] S.Jha, R.Yadava, IEEE Sensors Journal, 11(1) 35-44(2011) ; doi : 10.1109/JSEN.2010.2049351

[4] M.Penza, G.Cassaon. Sensors and Actuators B 89, 269-284 (2003); doi: 10.1016/S0925-4005(03)00002-9

[5] A. Hyvärinen, E.Ojar. Neural Networks 13, 411-430 (2000); doi: 10.1016/S0893-6080(00)00026-5

[6] M.Han, Y. Sun, Y.Fan. Expert Systems with Applications 34, 2905-2920 (2008); doi: 10.1016/j.eswa.2007.05.020

[7] C.Di. Natale, E.Martinellia. Sensors and Actuators B 82, 158-165 (2002); doi: 10.1016/S0925-4005(01)01001-2