A study on concentration estimation of formaldehyde in four-component gas mixtures using neural networks

Jinqing Qi¹, Jing Wang¹, Honglin Zhou¹, Guoping Jiang¹, Min Ji²

¹ School of Electronic and Information Engineering, Dalian University of Technology
Dalian, 116023, P.R.China
² School of Chemistry Engineering, Dalian University of Technology
Dalian, 116023, P.R.China
Tel.: + 86 [411] 84706002 Fax: +86 [411] 84706706 E-mail: jinqing@dlut.edu.cn

Abstract

In this study, the feedforward neural networks (FNNs) were applied for concentration estimation of formaldehyde. Contrasting with previous studies on binary gas mixtures, four-component gas mixtures containing formaldehyde, ethanol, acetone and toluene were investigated in this paper. Experimental results showed that Levenberg-Marquardt algorithm was fast and efficient for quantitative analysis of multicomponent gas mixtures and a relative prediction error of 0.26% for formaldehyde was achieved.

Keywords: Gas mixtures, formaldehyde, neural networks

1. Introduction

Indoor air pollution (IAP) is a significant health issue nowadays. The U.S. Environmental Protection Agency (EPA) risk studies have consistently ranked IAP among the top five environmental health risks to the public [1]. IAP is caused by many sources within the building. For example, building materials, cleaning agents, copy machines and printers may all emit volatile organic compounds (VOCs) such as formaldehyde (HCHO), toluene, ethanol, acetone etc. Among various VOCs, formaldehyde is recognized as one of most important IAP because it is widely used in the manufacturing of building materials. Moreover, formaldehyde is toxic for human even if its concentration in room air is low [2-3]. World Health Organization indicates that the regulation value of formaldehyde concentration in room air is less than 0.1 mg/m³ [4].

Therefore, finding effective methods to monitor formaldehyde is demanded for indoor air quality measurement and control. In the past two-three years, several micro-structure formaldehyde gas sensors based on metal oxide semiconductor materials have been proposed in the literature [5-7]. It is well known that metal oxide semiconductor gas sensor usually has a high sensitivity to toxic or flammable gases, but still have limitations and challenges such as selectivity, long-term stability, and so on. Besides the research on preparation and doping of semiconductor metal oxides, neural network (NN) technique is another research focus on improving the performance of gas sensors [7].

Many studies have shown that 20-130 kinds of VOCs can be detected in a non-industrial indoor environment. According to the study in [8], the most abundant VOCs were ethanol, followed by acetone, toluene, limonene etc. Consequently, contrasting with the previous studies on binary gas mixtures [7, 9], estimating the concentration of formaldehyde in four-component gas mixtures, containing (1) formaldehyde as a primary component, (2) ethanol, acetone and toluene as interference gases, is our motivation for this study.
2. Experimental and methods

In the last decade, artificial neural networks (ANNs) have been commonly used for analyzing multisensor data and have shown their efficiency in classification and quantitative analysis of gas mixtures. In order to estimate the concentration of formaldehyde in four-component gas mixtures, eight saw-based gas sensors and a feedforward neural network with one hidden layer were employed in our test system. The schematic diagram of the whole testing system and data acquisition subsystem are showed in Fig. 1 and Fig. 2, respectively. Sensors and their related sensing materials will be described in following sub-section. The eight saw-based sensors were fabricated with two types of sensing materials (four sensors for each sensing material). Sensor responses of two types of materials to a gas mixture are shown in Fig. 3. Obviously, the sensor responses to the formaldehyde gas are not so ideal in the gas mixtures. Consequently, a feedforward neural network with one hidden layer was employed in our test system.

![Fig. 1. Schematic diagram of the testing system and architecture of neural network.](image1)

![Fig. 2. Data acquisition subsystem (DAQ), a) Block diagram of the DAQ, b) the printed circuit board of the DAQ.](image2)

![Fig. 3. Sensor responses to a gas mixture.](image3)

2.1. Sensor fabrication

Several saw-based formaldehyde sensors coated with different sensing materials, such as polyisobutylene, polyepichlorohydrin, ethyl cellulose, were investigated in this study. Spin-coating method was adopted in sensing film fabrication. The thickness of films \((d)\) can be calculated using following formulation:

\[
d = \frac{C \cdot V \cdot \rho_2}{\rho_1 \cdot S}
\]

where \(\rho_1, \rho_2\) indicate the density of sensing material and solution, respectively; \(C\) and \(V\) indicate the concentration and volume, respectively; \(S\) indicates the area of the sensing film.
Sensing film thickness is a critical factor for sensor response. In order to find the optimum thickness, sensor response comparison with different sensing film thickness were investigated in our study. As shown in Fig. 4, the sensitivity of sensor can be improved dramatically while sensing film thickness was increased. For example, the sensitivity of sensor at film thickness of 45 nm is only 0.56 Hz/ppm, whereas, the sensor sensitivity of 2.31 Hz/ppm can be achieved at the film thickness of 270 nm.

On the other hand, the sensor response time was also depended on the sensing film thickness. Figure 5 shows the sensor response times at different sensing film thicknesses. As shown in Fig. 5, sensor response time is increased with the increase of sensing film thickness. For example, the sensor response time at 270 nm film thickness is double longer than it at 180 nm film thickness. Therefore, 180 nm film thickness was adopted in our study while both sensor response sensitivity and sensor response time were taken into account.

![Fig. 4. Sensor response sensitivity comparison with different sensing film thicknesses.](image1)

![Fig. 5. Sensor response time at different sensing film thicknesses.](image2)

2.2. Architecture of neural network

The number of hidden layer nodes is a key problem for a feedforward neural network. As shown in Fig. 1, there are totally eight nodes in the input layer that connected with eight sensors, respectively, and four nodes in the output layer each corresponding to one of concentration values of ethanol, acetone, toluene and formaldehyde. The architecture of neural network in our test system is 8-X-4 accordingly, where “X” indicates the number of hidden layer nodes to be determined.

In order to find the suitable number of hidden layer nodes, 47 groups of sample data series based on the responses of the eight sensors to the gas mixtures at different concentration were captured, in other world, totally 376 data were used for training the neural networks. The system errors of neural networks with different number of hidden layer nodes that varied from 4 to 20 are shown in Figure 6. As a result, a hidden layer with 10 nodes was selected in our testing system.

2.3. Backpropagation algorithms

Backpropagation (BP) algorithm has been widely used for training the weights of feedforward neural networks. It works well for many problems. Standard BP (SBP) algorithm applies the steepest descent method to update the weights, it suffers from many drawbacks such as local minimum, slow convergence rate etc. As a result, many related algorithms have been proposed to address that problem [10-11]. In this study, the Levenberg-Marquardt BP (LMBP) algorithm, an improved version of BP with the fastest convergence speed, was utilized during the training of neural networks. The result of convergence on LMBP learning algorithm is shown in Fig. 7.

As for a comparison, learning methods of conjugate gradient and Gauss-Newton were also investigated in this study. The comparison of convergence rate of neural networks with different learning methods is shown in Table 1. As shown in Table 1, LMBP algorithm achieves the fastest convergence with the smallest iteration number of 39 and, however, the
standard BP algorithm is failed in training the network. Therefore, LMBP algorithm was finally employed in our test system.

![Fig. 6. System errors of neural networks with different hidden layer nodes.](image)

![Fig. 7 Result of convergence on LMBP learning](image)

**3. Results and discussion**

As mentioned in the previous section, the training set composed of 47 samples and the maximum epoch’s number was used as 30000 for all training algorithm. The goal was $10^{-5}$ for a mean squared error (MSE). The standard BP algorithm provided slow convergence and the obtained mean squared error was far from the aimed value. Levenberg-Marquardt algorithm provided faster convergence and the lower MSE than other algorithms tested.

In order to test appropriateness of training algorithms, the relative error is defined as follows.

$$RE = \left| \frac{C_{predicted} - C_{real}}{C_{real}} \right| \times 100\%$$

where $C_{real}$ is the real concentration of gas and $C_{predicted}$ is the predicted concentration of gas using neural network.

The predicted concentrations and relative errors using LMBP neural network are shown in Table 2. The average relative errors to formaldehyde, ethanol and toluene are smaller than that to acetone.

<table>
<thead>
<tr>
<th>No.</th>
<th>Formaldehyde</th>
<th>Ethanol</th>
<th>Acetone</th>
<th>Toluene</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>100.2 (0.02%)</td>
<td>340.5(0.15%)</td>
<td>249.7(0.12%)</td>
<td>90.3(0.33%)</td>
</tr>
<tr>
<td>2</td>
<td>119.9 (0.08%)</td>
<td>370.2(0.54%)</td>
<td>360.5(0.14%)</td>
<td>380.3(0.08%)</td>
</tr>
<tr>
<td>3</td>
<td>99.5 (0.05%)</td>
<td>290.6(0.2%)</td>
<td>89.8(1.33%)</td>
<td>260.1(0.04%)</td>
</tr>
<tr>
<td>4</td>
<td>59.9 (0.16%)</td>
<td>69.9(0.14%)</td>
<td>140.4(0.28%)</td>
<td>108.1(0.06%)</td>
</tr>
<tr>
<td>5</td>
<td>79.7 (0.37%)</td>
<td>400.5(0.12%)</td>
<td>260.1(0.04%)</td>
<td>280.2(0.07%)</td>
</tr>
<tr>
<td>Average RE(%)</td>
<td>0.26%</td>
<td>0.67%</td>
<td>1.9%</td>
<td>0.11%</td>
</tr>
</tbody>
</table>

Table 1. Comparison of convergence rate.

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Iteration numbers</th>
</tr>
</thead>
<tbody>
<tr>
<td>SBP</td>
<td>30000 (failed)</td>
</tr>
<tr>
<td>CGBP</td>
<td>1117</td>
</tr>
<tr>
<td>GNBP</td>
<td>467</td>
</tr>
<tr>
<td>LMBP</td>
<td>39</td>
</tr>
</tbody>
</table>

Table 2. Predicted concentrations and relative error.
4. Conclusion
In this study, a three-layers backpropagation neural network was applied for concentration estimation of formaldehyde in four-component gas mixtures. Experimental results showed that Levenberg-Marquardt algorithm was fast and efficient for quantitative analysis of multicomponent gas mixtures and a relative prediction error of 0.26% for formaldehyde was achieved.

5. Acknowledgment
The authors thank to the financial support from the National Natural Science Foundation of China (60474052).

References