

基于 PCA-BP 神经网络对甲醛和甲醇的识别研究

宋海声, 麻林召*, 王一帆, 朱恩功, 李承飞

西北师范大学物理与电子工程学院, 甘肃 兰州 730070

摘要 当前电子鼻对有毒气体的识别存在数据量少, 训练生成的神经网络映射能力差等问题。本文以甲醛和甲醇为目标气体, 采用自制的气敏传感器对甲醛和甲醇进行数据采集, 并对采集到的数据进行滤波和平滑处理, 以提取不同传感器对目标气体的响应值。依据准则函数生成伪随机数, 并建立伪随机特征值矩阵以扩大有效数据量。利用主成分分析 (PCA) 法对特征值进行降维处理, 选择贡献率大的主元成分作为反向传播 (BP) 神经网络的输入向量, 构造 PCA-BP 神经网络。分别用实测特征值矩阵和伪随机特征值矩阵训练 PCA-BP 神经网络, 通过对比分析两个网络得出, 实测特征值矩阵的识别率为 92%, 而伪随机特征值矩阵的识别率为 97%。结果表明, 伪随机特征值矩阵能有效提高 PCA-BP 神经网络的映射能力, 提高识别正确率。

关键词 测量; 模式识别; 伪随机特征值矩阵; 反向传播神经网络; 主成分分析法

中图分类号 X831

文献标志码 A

doi: 10.3788/LOP57.071201

Recognition of Formaldehyde, Methanol Based on PCA-BP Neural Network

Song Haisheng, Ma Linzhao*, Wang Yifan, Zhu Engong, Li Chengfei

College of Physics and Electronic Engineering, Northwest Normal University, Lanzhou, Gansu 730070, China

Abstract At present, the identification of toxic gases by electronic noses has a small amount of data, and the ability of neural network mapping generated by training is insufficient. In this work, the formaldehyde and methanol are used as target gases, and collected by self-made gas sensor. After filtering and smoothing the collected data, the different response values are extracted. The pseudo-random numbers are generated according to the criterion function, and the pseudo-random matrix is established to expand the effective data volume. The principal component analysis (PCA) is used to reduce the dimensionality of the eigenvalues, and the principal component score with large contribution rate is selected as the input vector of the back-propagation (BP) neural network to construct PCA-BP neural network, which is trained by using the measured eigenvalue matrix and the pseudo-random eigenvalue matrix respectively. By comparing the two networks, the recognition rate of the measured eigenvalue matrix is 92%, and the recognition rate of the pseudo-random eigenvalue matrix is 97%. The results show that the pseudo-random eigenvalue matrix can effectively improve the mapping ability of BP neural network and the accuracy of recognition.

Key words measurement; pattern recognition; pseudo-random eigenvalue matrix; back propagation neural network; principal component analysis

OCIS codes 120.0280; 100.4996; 150.1135

1 引言

甲醛和甲醇是日常生活中常见的有毒物质。甲醛对皮肤和黏膜具有强烈的刺激作用, 是引发鼻咽癌、鼻窦癌、血癌和肺癌的主要因素之一。甲醇可以

经过皮肤和呼吸道进入人体, 刺激人的神经系统, 具有麻痹效果, 同时对视网膜和视觉神经具有选择性, 可以导致视觉神经损伤, 严重可能导致失明。因此对甲醛和甲醇进行识别具有十分重要的意义。对甲醛的识别, 常见的方法有色谱法^[1]、气相色谱——质

收稿日期: 2019-10-09; 修回日期: 2019-11-07; 录用日期: 2019-11-26

基金项目: 国家自然科学基金(11664036)、甘肃省自然科学基金(1606RJZA065)

* E-mail: 1093704655@qq.com

谱(GC-MS)联用法^[2];对甲醇的识别,常见的方法有顶空进样-气相色谱法^[3]、傅里叶变换红外法^[4]。这些物理方法在一定程度上可以对甲醛和甲醇进行识别,但是操作复杂,需要专业的人士才能进行操作,不利于在现实生活中推广和使用。

随着传感器技术的迅猛发展。电子鼻在环境监测、食品安全及药物识别等方面发挥了极大的作用。郑豪男等^[5]基于微纳米传感器探究了有毒有害气体检测的方法,对于每种浓度的气体重复测量5次。钱小瑞等^[6]应用反向传播(BP)神经网络对室内甲醛进行了定量分析,结果表明,BP神经网络可以很好地对微量甲醛浓度进行定量预测。张覃轶等^[7]基于ZnO气敏传感器阵列对乙醇、丙酮、苯、甲苯、二甲苯等目标气体进行了相关的识别研究,对每种气体浓度重复测量13次。以上研究的目标气体具有毒性,数据重复测量的次数少,造成采集的原始数据量较少,识别效果欠佳。何爱香等^[8]基于温度调制的气体传感器获取探测的动态信号,每种调制波形重复测量了96次,结合主成分分析对数据进行降维处理,对三种气体进行了有效的识别。王雨等^[9]基于流速调制的电子鼻对啤酒进行分类,每个样本重复测量25次,每次需要300s的时间,识别的正确率为96.8%。因为对气体识别需要进行大量的实验,以获取较多的实验数据,用来提高识别的正确

率。同时长时间地进行测量,在一定程度上占用了大量的时间和资源,不同样本特性的提取存在数据量不足等问题。为了降低实验操作成本,并能够有效地对目标气体进行识别,本文通过建立伪随机特征值矩阵克服这些缺点。伪随机特征值矩阵的具体建立过程如下:首先采用实测数据响应曲线,提取同一种传感器的最大响应值作为特征值^[10],并将特征值投影到坐标轴上;其次通过求平均值方法找到不同数据集的中心;最后用准则函数对数据集中心的数值进行提取,形成新的特征值,并构造成伪随机特征值矩阵。采用主成分分析(PCA)法将新生成的伪随机特征值矩阵进行降维处理,将贡献率较大的主元作为BP神经网络的输入。本文以甲醛和甲醇的响应值数据为基础样本,利用生成的伪随机特征值矩阵对甲醛和甲醇进行分类识别。

2 实验分析与数据处理

2.1 实验分析

实验设备如图1所示,1号设备为数据采集卡,一端和电脑连接,另一端和信号调理电路连接;2号设备为直流电源;3号设备为气体传感器阵列;4号设备为电风扇;5号设备为排气盖;6号设备为测试腔;7号设备为注射器;8号为待测液体;9号为加热板。

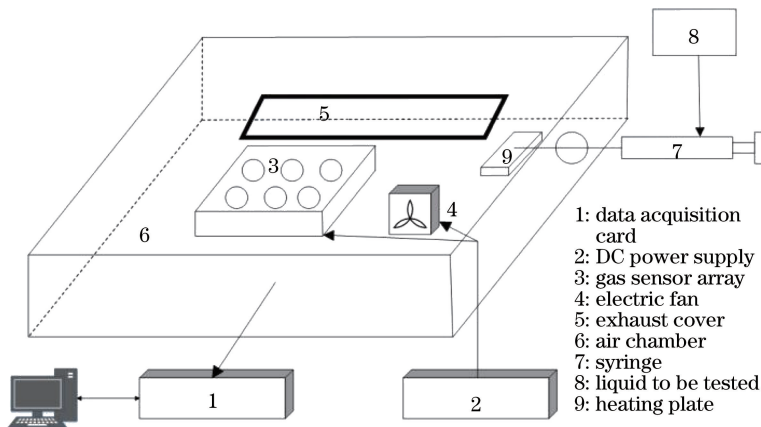


图1 实验系统框图

Fig. 1 Experimental system block diagram

实验中使用实验室自制的传感器,制备过程描述如下:1)采用水热法^[11]分别制取了 $\text{SnO}_2/\text{Pr}^{3+}$, PrFeO_3 , $\text{PrFeO}_3/\text{Sm}^{3+}$, BiVO_4 , $\text{SnO}_2/\text{In}_2\text{O}_3$, SnO_2/Pa , SnO_2/Sb 7种半导体纳米材料;2)用去离子水将以上7种半导体纳米材料调配成均匀的糊状,并均匀地涂在带加热丝的陶瓷管上,厚度大约为 $300\ \mu\text{m}$;3)将涂抹后的陶瓷管干燥后置于老化台,

老化10天时间,使其性能更加稳定。实验中测试腔的温度为 $23\ ^\circ\text{C}$,测量过程为:1)对测试腔进行干燥洗气处理,排除其他气体对测量气体产生的测量误差,干燥和洗气时间为5min;2)对传感器阵列进行加热,使其传感器的波形基线稳定,减少测量误差,加热时间为2min;3)用微量注射器抽取一定体积的甲醛和甲醇分别注射到加热板上,使其

快速挥发;4)气体完全挥发后,采集 1 min 气体浓度的响应值;5)1 min 后对测试腔进行放气,直到波形恢复到基线位置。对甲醛和甲醇三种浓度(2 μL , 4 μL , 6 μL)重复测量 30 次,利用气体传感器阵列采集甲醛和甲醇的响应值,通过数据处理获得各个传感器的响应值曲线。使用 BP 神经网络

对甲醛和甲醇这两种有机物进行分类识别,具体的步骤如图 2 所示。1)将气体通入气体传感器阵列;2)信号测量电路采集气体浓度产生的电压值;3)采集到的电压值通过信号调理电路对信号进行预处理,送入计算机;4)对预处理的数据通过误差神经网络,进行识别分析。

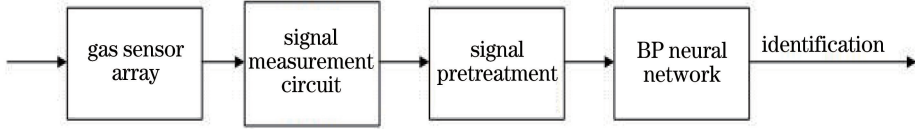


图 2 气体识别过程

Fig. 2 Gas identification process

2.2 数据处理

采用液体定量取样方法,即微量注射器每次抽取定量的(2 μL , 4 μL , 6 μL)两种液体注射到测试腔的加热板进行加热,并进行测试,记录各个传感器的响应值(实验中每组数据循环测量 30 次)。由于信号采集过程中,电阻元件会受到温度的影

响,导致采集到的模拟信号经过模数转换(ADC)时引入了高频噪声信号。如果将采集的信号直接送入计算机造成识别效果不佳,因此首先要对原始信号进行预处理。采用均值滤波,对波形进行平滑处理。图 3 所示为选取 4 μL 甲醇溶液的响应信号进行预处理对比。

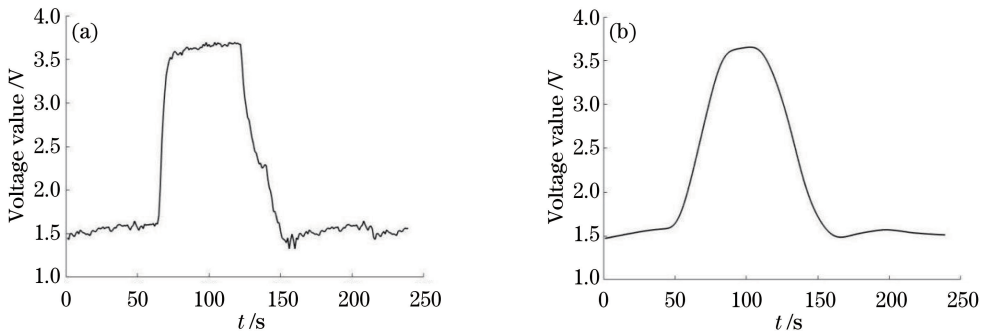


图 3 采集数据波形。(a)原始数据波形;(b)滤波后的数据波形

Fig. 3 Acquisition of data waveform. (a) Raw data waveform; (b) filtered data waveform

实验中采用自制的传感器,因此需要对数据的稳定性进行分析,传感器响应值的相对标准差可以反映传感器的稳定程度。传感器对被测气体反应越稳定,其相对标准差就越小,计算公式为

$$\delta = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - \eta)^2} / \eta, \quad (1)$$

式中: N 为样本数; η 为样本均值; x_i 为样本值。

根据(1)式对采集的数据进行统计分析,表 1 列出了制备 7 种传感器响应值的相对标准差。从表 1 可以看出,7 种气敏传感器对甲醛和甲醇在不同浓度下响应值的相对标准差均较小,因此满足传感器阵列稳定性要求。通过对比发现,7 种气敏传感器采集甲醛和甲醇的响应值有较大差异,因此可以通过分析响应值,对甲醛和甲醇这两种有机物进行识别。

3 算法分析

3.1 BP 神经网络

模式识别是指将实验采集的气体数据样本通过合适的处理方法得到不同识别结果的过程。其中,误差反向传播算法^[12]是最常用的模式识别算法。BP 神经网络将输出期望值与测量值的差值进行反向传播,修改每层神经元的权值和阈值,直到满足网络设定的期望误差。BP 神经网络具有较好的非线性映射能力和网络结构简单的优势,如网络的各层数和神经元数可以根据数据量大小进行设定。通常 3 层结构的 BP 神经网络可以满足对任意非线性函数进行逼近的要求。常见的 BP 神经网络包含网络输入层、隐含层和输出层。

表1 气体传感器响应值的相对标准差

Table 1 Relative standard deviation of gas sensor response

Sensor		Formaldehyde / μL			Methanol / μL		
		2	4	6	2	4	6
1	σ	0.0666	0.0621	0.0441	0.0031	0.0027	0.0021
	μ	3.2691	3.6191	4.9390	3.9663	4.3863	5.9063
2	σ	0.0116	0.0102	0.0074	0.0048	0.0036	0.0027
	μ	3.9011	4.4411	6.1211	3.5445	4.7445	6.3245
3	σ	0.0071	0.0066	0.0066	0.0039	0.0025	0.0018
	μ	4.5214	4.8714	6.0893	2.8882	4.5382	6.3982
4	σ	0.0115	0.0096	0.0096	0.0056	0.0041	0.0031
	μ	3.8311	4.5811	6.6411	3.1325	4.3825	5.8625
5	σ	0.0112	0.0087	0.0074	0.0099	0.0086	0.0064
	μ	3.5972	4.5972	5.4572	3.7638	4.3438	5.8338
6	σ	0.0161	0.0131	0.0107	0.1048	0.0669	0.0669
	μ	3.4628	4.3128	5.2388	0.9534	1.4934	2.0161
7	σ	0.0261	0.0221	0.0165	0.3765	0.3369	0.3369
	μ	3.4741	4.1141	5.4641	3.0595	3.1495	5.1293

3.2 PCA-BP 神经网络模型

主成分分析算法^[13]是对数据进行处理的常用方法之一,通过对数据集进行线性组合求解得出协方差矩阵的最大 k 个特征值对应的特征向量,原数据集有 n 个特征值,在尽量减少对原始数据信息丢失的情况下对数据进行降维处理。在 k 个特征值中选择方差最大的 p_1 为第一主成分,若 p_1 的贡献率较小,不足以代表原来 n 个特征值代表的信息,则选取 p_2 即第二主成分,若 p_1 和 p_2 两个主成分的累积贡献率不能满足实际需要,继续选取 p_3 、 p_4 等直到满足实际需要为止,且 p_1, p_2, \dots, p_n 各不相关。因此,可以构造 T 个主成分,构造公式为

$$\begin{cases} p_1 = l_{11}x_{11} + l_{12}x_{12} + \dots + l_{1n}x_{1n} \\ p_2 = l_{21}x_{21} + l_{22}x_{22} + \dots + l_{2n}x_{2n} \\ \vdots \\ p_n = l_{n1}x_{n1} + l_{n2}x_{n2} + \dots + l_{nn}x_{nn} \end{cases}, (2)$$

式中: $l_{1j}^2 + l_{2j}^2 + \dots + l_{nj}^2 = 1, (j = 1, 2, \dots, n)$, x_{mj} 为特征值归一化数值,从 n 个变量中选取 T 个变量 ($T < n$) 代替原始数据的大部分信息,其余 $n - T$ 个变量对原始信息的贡献率太小,因此选取 T 个变量作为 BP 神经网络的输入。

BP 神经网络基于梯度下降法,网络收敛速度较慢,误差函数容易陷入局部最小值,当对原始数据提取过多的特征向量时,向量彼此间存在线性相关,当特征向量选取过少时,网络训练结果不佳,因此在选择网络输入时,特征值维数不宜选取过多,否则会使网络过度拟合导致识别率下降,同时浪费大量的时间对网络进行训练,建模时间过长。针对以上存在

的问题,建立 PCA-BP 神经网络模型,对采集到的原始数据进行特征值提取,通过 PCA 法对特征值进行降维处理,得到的主成分彼此不相关,选取贡献率大的主元成分作为网络的输入。利用 PCA 法对数据进行降维处理,减少数据间的相关性,同时降低网络的输入向量的维数,缩短网络建模的时间,通过精简数据维数,克服网络过度拟合,提高网络分类识别的正确率。PCA-BP 神经网络结构图如图 4 所示。

3.3 伪随机特征值矩阵算法

1) 当分析对象有 n 个特征值时,这些特征值构成 n 维特征矢量 $\mathbf{x} = (x_1, x_2, \dots, x_n)$, 其中 $i = 1, 2, \dots, n$ 。对于 m 个分析对象的特征值均表示成 n 维特征矢量,则这 m 个分析对象对应测量值组合在一起形成 $m \times n$ 维测量矩阵 $\mathbf{A}_{m \times n}$, 矩阵元素为 x_{mi} 。

2) 矩阵 $\mathbf{A}_{m \times n}$ 的每一列向量 $\mathbf{x} = (x_{1i}, x_{2i}, \dots, x_{mi})^T, i = 1, 2, \dots, n$, 将其投影到 x 轴上,求得这些数据的中心点为

$$\bar{x} = \frac{x_{1i} + x_{2i} + \dots + x_{mi}}{m}, i = 1, 2, \dots, n. \quad (3)$$

3) 数据集中心点间的距离为

$$d = \max \left| \sum_{i=1}^m x_i - \bar{x} \right|. \quad (4)$$

4) 设伪随机特征值矩阵为 $\mathbf{W}_{m \times n}$, 形式如下

$$\mathbf{W} = \begin{bmatrix} \tau_{11} & \tau_{12} & \dots & \tau_{1n} \\ \tau_{21} & \tau_{22} & \dots & \tau_{2n} \\ \vdots & \vdots & & \vdots \\ \tau_{m1} & \tau_{m2} & \dots & \tau_{mn} \end{bmatrix}_{m \times n}, \quad (5)$$

式中: τ_{mn} 是在 x_{mn} 左右进行波动。

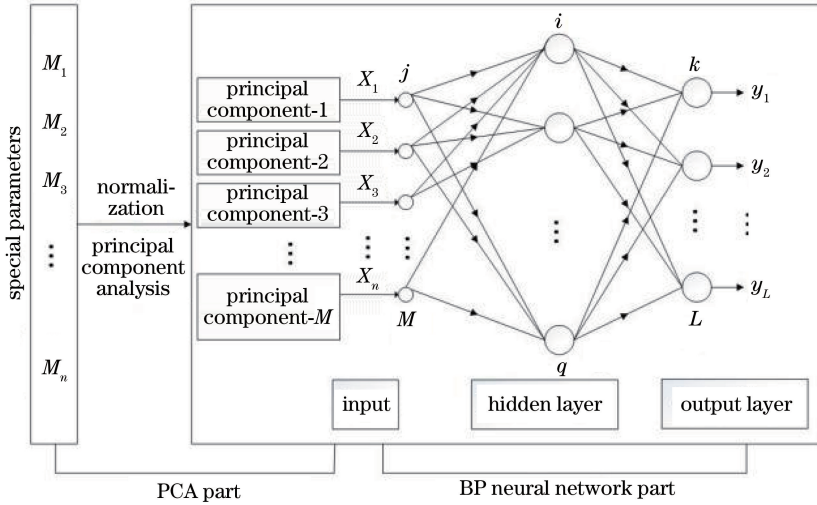


图 4 PCA-BP 神经网络模型

Fig. 4 PCA-BP neural network model

5) 矩阵 $W_{m \times n}$ 中的元素满足如下准则函数

$$\tau_{mn} = \begin{cases} \bar{x} + \varepsilon d, \bar{x} \leq \tau_{mn} \leq \bar{x} + d \\ \bar{x} - \varepsilon d, \bar{x} - d \leq \tau_{mn} \leq \bar{x} \end{cases}, \quad (6)$$

式中: ε 的取值为 $(-1.5, 1.5)$ 。

6) 矩阵 $W_{m \times n}$ 的行向量分别为 $\tau = (\tau_{i1}, \tau_{i2}, \dots, \tau_{in})$, $\mu = (\mu_{j1}, \mu_{j2}, \dots, \mu_{jn})$, 其中 $i, j = (1, 2, \dots, m, i \neq j)$ 。通过计算矩阵的相关系数, 对组成的特征向量进行相似度分析, 其公式为

$$\rho = \frac{\text{cov}(\tau, \mu)}{\sqrt{\text{cov}(\tau, \tau) \cdot \text{cov}(\mu, \mu)}}, \quad (7)$$

式中: $\text{cov}(\tau, \mu)$ 为两组信号的协方差, 相关系数 ρ 越接近 0, 新构造的行向量 τ 和 μ 彼此间的相似程度越小, 数据对分类识别价值更大。

3.4 特征值提取与伪随机矩阵

由于传感器阵列采集到的数据量较大, 直接进行分类识别效果较差。为解决这一问题, 通常采用提取特征曲线的特征值进行分析, 在这一过程中, 特征参数的选取对模式识别起着关键作用。因此, 在特征参数选择时, 应使响应曲线的数据趋于稳定, 使

特征参数包含大多数传感器最有价值的特征响应信号。在实验测量当中, 选取传感器每条响应曲线稳定后的最大值作为提取的特征值。

实验中传感器阵列有 7 种气敏传感器, 因此有 7 个提取的特征值。实验当中每组数据循环测量 30 次, 因此得到矩阵 A 为 60×7 维特征值矩阵。伪随机特征值矩阵 W 的构造步骤为: 1) 分别取出矩阵 A 的列向量, 并求得每列向量对应的数据中心; 2) 再次求得数据中心与列向量值的最远距离; 3) 通过准则函数在中心点附近随机取值; 4) 按照每列向量对应的属性进行排列, 并形成新的矩阵。通过步骤 3 随机生成 100 组特征值, 通过步骤 4 组成矩阵 W 为 200×7 维矩阵。对于新组成的矩阵 W 行向量进行相似度分析。表 2 为 7 个传感器部分数据相似度标征, 表中 τ_i 为 W 的行向量。由表 2 可以看出, W 行向量为相似度的值, 且主对角线的值均为 1, 因为是同一个向量的相关系数。其他向量的相关系数都接近 0, 可以认为向量间的相似程度很小, 满足计算要求。

表 2 部分 W 矩阵数据相似度标征

Table 2 Data similarity of partial W matrix

τ_i	1	2	3	4	5	6	7
1	1	0.2455	-0.3589	-0.4109	0.1243	0.1406	0.3054
2	0.2455	1	-0.0841	0.0651	-0.4414	0.1488	0.5628
3	-0.3589	-0.0841	1	0.2956	-0.2973	0.0211	-0.3623
4	-0.4109	0.0651	0.2956	1	0.0373	-0.1307	0.0462
5	0.1243	-0.4414	-0.2973	0.0373	1	-0.0602	0.0502
6	0.1406	0.1488	0.0211	-0.1307	-0.0602	1	0.0583
7	0.3054	0.5628	-0.3623	-0.0462	0.0502	0.0583	1

3.5 特征值主成分分析

由表 2 可以看出,生成的特征向量中存在一定相关性,可以通过 PCA 法进行降维处理,通过对较多的特征向量进行线性组合,重新生成互相独立的、同时包含原始数据大部分信息的新特征向量。从 W 矩阵中随机抽取甲醛和甲醇各 20 个样本进行主成分分析,前 7 个主成分的贡献率依次为 59.69%、40.06%、0.186%、0.037%、0.0146%、0.002%、0.001%,累积贡献率为 59.69%、99.75%、99.93%、99.97%、99.98%、99.99%、

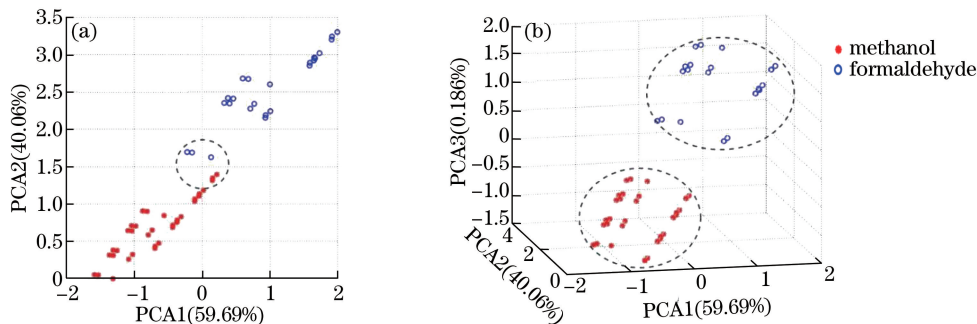


图 5 PCA 得分图。(a)前 2 个主成分;(b)前 3 个主成分

Fig. 5 PCA score chart. (a) First two main components; (b) first three main components

4 PCA-BP 神经网络对甲醛和甲醇的分类结果

BP 神经网络可以通过已知样本对未知样本进行分类。将响应曲线的特征值作为系统输入,通过对神经网络进行训练,并不断地调整权值和阈值,使输入和输出建立近似的非线性关系,从而达到精确分类的目的。

实验中采用 7 个传感器进行数据采集,从每条响应曲线中提取 1 个最大值作为特征值,共有 7 个特征向量。对 7 维特征向量进行降维,采用前 3 个主成分作为 BP 神经网络的输入量。实验需要识别甲醛和甲醇两种有机物,因此 BP 神经网络的输入层、隐藏层、输出层的神经元个数分别为 3、15、2,每个输出元对应一种液体的类别,其网络拓扑结构为 $3 \times 15 \times 2$ 。本实验对甲醛和甲醇两种有机物在同一浓度下进行重复测量,共获得 60 个样本,提取响应曲线的特征值,并组成特征值矩阵 A 。当采用 BP 神经网络进行分析时,矩阵 A 中训练样本选自每种样品的 10 个样本值,测试样本选自每种样品的 20 个样本值,即两种有机物的训练样本为 20 个,测试样本为 40 个。由矩阵 A 构造生成的矩阵 W 中,训练样本选自每种样品的 50 个样本值,测试样本选自每种样品的 50 个样本值,即训练样本值 100 个,测试样本

100%。前 2 个主成分累积贡献率为 99.75%,包含数据大部分信息。图 5 为 PCA 得分图分布情况,根据前 2 个主成分得分图 5(a)可知,用线圈起来的部分数据彼此间距离过近,会导致后面的分类中误差较大。图 5(b)是对前 3 个主成分进行分析,可以很容易地将甲醛和甲醇进行分类,在这之后,随着主成分数量的增加,累积贡献率的增加相当缓慢。因此,可以利用前 3 个主成分代替 7 个传感器采集到的数据进行分类,进而达到降维的目的。

值 100 个。利用这两个特征值矩阵分别对 PCA-BP 神经网络进行训练和测试,对应甲醛和甲醇的网络输出设置为 -1 和 1。训练过程中不断改变神经元的权值,当满足误差精度时,结束训练。用矩阵 A 训练的 PCA-BP 神经网络记为 A -PCA-BP,用矩阵 W 训练的 PCA-BP 神经网络记为 W -PCA-BP。实验中选取 3 层神经网络,其中隐含层神经元数为 15,同时以 tansig 函数作为隐含层神经元的激励函数,将 purelin 函数作为输出层神经元的激励函数,训练函数采用共轭梯度法(trainscg),网络最大训练步长设置为 2000,目标误差定为 0.01,使用的软件为 Matlab2016a 版本。两个 BP 神经网络的分类结果如图 6 所示。

为了验证由矩阵 W 训练得到的 W -PCA-BP 网络泛化效果。将实测矩阵 A 用 W -PCA-BP 网络进行分类识别。识别误差率(MSE)定义为 BP 神经网络对错误区分的总样本数和测试样本数的比值。甲醛和甲醇的 MSE 由下式计算,

$$E_{1\text{MSE}} = \left| \frac{N_1 - N_{1\text{-BP}}}{N_1} \right| \times 100\%, \quad (8)$$

$$E_{2\text{MSE}} = \left| \frac{N_2 - N_{2\text{-BP}}}{N_2} \right| \times 100\%, \quad (9)$$

式中: N_1 和 N_2 分别为甲醛和甲醇的测试样本; $N_{1\text{-BP}}$ 和 $N_{2\text{-BP}}$ 分别为通过 BP 神经网络对甲醛和甲醇正确分类的样本。将实测矩阵 A 用 W -PCA-BP 网络进行分类识别的 MSE 结果列于表 3 中。从表

3 可以看出,通过矩阵 \mathbf{W} 训练的 PCA-BP 神经网络

对实测矩阵 \mathbf{A} 在识别上有了很大的提高。

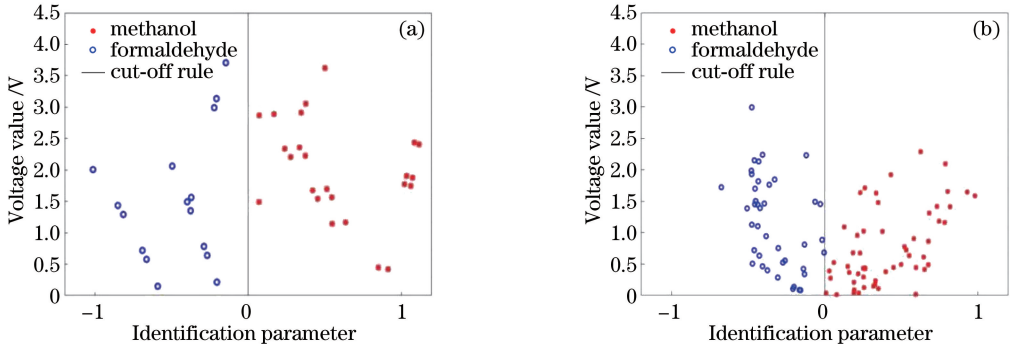


图 6 BP 神经网络分类结果。(a) \mathbf{A} -PCA-BP 分类;(b) \mathbf{W} -PCA-BP 分类

Fig. 6 Classification results of BP neural network. (a) \mathbf{A} -PCA-BP classification; (b) \mathbf{W} -PCA-BP classification

表 3 \mathbf{W} -PCA-BP 网络对矩阵 \mathbf{A} 识别结果

Table 3 Recognition results of matrix \mathbf{A} on

\mathbf{W} -PCA-BP network

Sample type	Study samples	Recognition result /piece	Identification error /%
Formaldehyde	30	31	3.3
Methanol	30	29	3.3

将矩阵 \mathbf{A} 和矩阵 \mathbf{W} 的数据分别用矩阵 \mathbf{A} 训练

的神经网络和矩阵 \mathbf{W} 训练的神经网络对甲醛和甲醇进行识别对比,结果列于表 4 中。从表 4 可以看出,用矩阵 \mathbf{A} 训练的神经网络,对识别的精度存在较大的误差,从表 5 可以看出,用矩阵 \mathbf{W} 训练的神经网络在识别精度上有了很大的提高,但是所用的时间较长,这是由于每次训练的数据增加导致。所以在处理数据量较大时,可以用 PCA 进行降维处理,不仅节省时间还能提高精度。

表 4 矩阵 \mathbf{A} 训练的网络对甲醛和甲醇的识别

Table 4 Identification of formaldehyde and methanol by matrix \mathbf{A} trained network

Method	Formaldehyde identification samples	Methanol identification samples	Total number of samples	Number of error /piece	Processing time /s
\mathbf{A} -BP	38	22	60	8	3.2
\mathbf{W} -BP	100	100	200	8	3.5
\mathbf{A} -PCA-BP	35	25	60	5	3.0
\mathbf{W} -PCA-BP	94	106	200	6	3.3

表 5 矩阵 \mathbf{W} 训练的网络对甲醛和甲醇的识别

Table 5 Identification of formaldehyde and methanol by matrix \mathbf{W} trained network

Method	Formaldehyde identification samples	Methanol identification samples	Total number of samples	Number of error /piece	Processing time /s
\mathbf{A} -BP	27	33	60	3	4.5
\mathbf{W} -BP	102	98	100	2	5.0
\mathbf{A} -PCA-BP	31	29	60	1	2.8
\mathbf{W} -PCA-BP	95	105	200	5	3.1

5 结 论

本文利用自制传感器采集甲醛和甲醇的气敏响应值,根据数学方法对采集的数据进行了有效的扩增,建立了伪随机数矩阵。通过该伪随机数对 BP 神经网络进行训练,有效地提升了网络的泛化能力,使网络模型对训练样本以外的数据具有正确的映射能力。通过 PCA 方法有效降低了数据的维数,建立了合适的伪随机特征值矩阵,有效地对甲醛和甲醇

进行识别分类。该研究有助于对不宜长时间采集,或者数据获取难度大的有毒气体进行识别分类,可以采取该方法对数据量进行扩充,有效提升 PCA-BP 神经网络识别的正确率,同时节省实验的周期。

参 考 文 献

[1] Peng L Z, Chen W, Liu H, et al. Comparison of the determination methods of formaldehyde value recommended by the national standards [J]. China

- Standardization, 2019(11): 133-135, 157.
- 彭力争, 陈威, 刘晖, 等. 国标甲醛量值识别方法比较[J]. 中国标准化, 2019(11): 133-135, 157.
- [2] Huang S H, Han Y, Xu P H, et al. Determination of methanol and formaldehyde in ophthalmic perfluoropropane gas by GC-MS [J]. Chemical Analysis and Meterage, 2019, 28(1): 38-41.
黄书浩, 韩银, 徐萍华, 等. 气相色谱-质谱法测定眼用全氟丙烷气体中甲醇、甲醛[J]. 化学分析计量, 2019, 28(1): 38-41.
- [3] Wang L, Li J C, Qiao L Y, et al. Determination of methanol and ethanol in ShuangWuZhiTong tincture by headspace sampling GC[J]. Strait Pharmaceutical Journal, 2019, 31(8): 106-108.
王璐, 李金慈, 乔立业, 等. 顶空进样-气相色谱法测定双乌止痛酊中甲醇和乙醇含量[J]. 海峡药学, 2019, 31(8): 106-108.
- [4] Luo Z F. Determination of methanol content in cosmetics by chemical method [J]. Chemical Engineering Design Communications, 2019, 45(6): 8-9.
罗志烽. 化学法测定化妆品中甲醇含量的方法[J]. 化工设计通讯, 2019, 45(6): 8-9.
- [5] Zheng H N, Chen Z Z, Shi P Y, et al. Study of toxic gases detection based on miniaturized sensor [J]. Chinese Journal of Sensors and Actuators, 2019, 32(4): 514-519.
郑豪男, 陈珍珍, 施佩影, 等. 基于微纳传感器的有毒有害气体检测方法研究[J]. 传感技术学报, 2019, 32(4): 514-519.
- [6] Qian X R, Wu F. Application of BP neural network in quantitative analysis of indoor formaldehyde [J]. Transducer and Microsystem Technologies, 2018, 37(4): 151-154.
钱小瑞, 吴飞. BP神经网络在室内甲醛定量分析中的应用[J]. 传感器与微系统, 2018, 37(4): 151-154.
- [7] Zhang Q Y, Xie C S, Li D F, et al. Recognition of ethanol, acetone, benzene, toluene and xylene using nano ZnO gas sensor array [J]. Chinese Journal of Sensors and Actuators, 2006, 19(3): 552-554, 558.
张覃轶, 谢长生, 李登峰, 等. 基于纳米 ZnO 气体传感器阵列的乙醇、丙酮、苯、甲苯、二甲苯的识别研究[J]. 传感技术学报, 2006, 19(3): 552-554, 558.
- [8] He A X, Tang Z A, Wei G F, et al. Study on dynamic signal recognition of gas sensor based on temperature modulation [J]. Transducer and Microsystem Technologies, 2015, 34(6): 24-26.
何爱香, 唐祯安, 魏广芬, 等. 基于温度调制的气体传感器动态信号识别研究[J]. 传感器与微系统, 2015, 34(6): 24-26.
- [9] Wang Y, Xing J G, Fu J, et al. Electronic nose based on flow velocity modulation and its application in beer classification [J]. Transducer and Microsystem Technologies, 2018, 37(11): 158-160.
王雨, 邢建国, 傅均, 等. 基于流速调制的电子鼻及其在啤酒分类中的应用[J]. 传感器与微系统, 2018, 37(11): 158-160.
- [10] Liu J X, Du B, Deng Y Q, et al. Terahertz-spectral identification of organic compounds based on differential PCA-SVM method [J]. Chinese Journal of Lasers, 2019, 46(6): 0614039.
刘俊秀, 杜彬, 邓玉强, 等. 基于差分-主成分分析-支持向量机的有机化合物太赫兹吸收光谱识别方法[J]. 中国激光, 2019, 46(6): 0614039.
- [11] Zhao L Y, Xi X L, Fan Y S, et al. Review of recent progress in the synthesis of nano-tungsten oxide via hydrothermal/solvothermal method and the application [J]. Materials Review, 2019, 33(19): 3203-3209.
赵林艳, 席晓丽, 樊佑书, 等. 纳米氧化钨的水热/溶剂热法制备及应用的综述[J]. 材料导报, 2019, 33(19): 3203-3209.
- [12] Umut Hasan, Mamat Sawut, Ma C Y . Hyperspectral estimation of wheat leaf water content using fractional differentials and successive projection algorithm-back propagation neural network [J]. Laser & Optoelectronics Progress, 2019, 56(15): 153002.
吾木提·艾山江, 买买提·沙吾提, 马春玥. 基于分数阶微分和连续投影算法-反向传播神经网络的小麦叶片含水量高光谱估算[J]. 激光与光电子学进展, 2019, 56(15): 153002.
- [13] Xu X J, Wang X S, Li A Z, et al. Fast classification of tea varieties based on laser-induced breakdown spectroscopy [J]. Chinese Journal of Lasers, 2019, 46(3): 0311003.
徐向君, 王宪双, 李昂泽, 等. 基于激光诱导击穿光谱的茶叶品种快速分类[J]. 中国激光, 2019, 46(3): 0311003.