

# 三元气敏阵列和有机溶剂识别\*

林海安 吴冲若 马骏

(东南大学电子工程系 南京 210018)

**摘要** 本文用  $\text{SnO}_2$  气敏元件构造了三元阵列,使用  $k$ -最近邻和加权  $k$ -最近邻两种分类器。结果表明,前一分类器对乙醇和汽油的识别率分别为 40% 和 90%,后一类分类器则分别为 70% 和 90%。模式识别是改善气敏元件选择性的有效途径。

**关键词** 气敏元件,模式识别,阵列,最近邻

## 1 引言

气敏传感器的发展有助于改善化学分析需要复杂、昂贵设备的现状。然而目前大多数气敏元件的选择性都比较差,表现出非线性的交叉灵敏度。为了解决这个问题, Persand 等<sup>[1]</sup>于 1982 年提出了气敏元件阵列和模式识别的方法,其实质是用发展成熟的信息处理技术来补偿相对发展缓慢的信息获得技术。经过 10 余年的努力,这一领域已经取得了一系列有价值的成就。就阵列构造而言,虽然目前报道的大多数阵列仍然是由分立元件组合而成的,但阵列的集成化已成为必然的趋势。在模式识别方法方面,主要有两类:一为多变量分析法,一为神经网络法。

气敏元件阵列和模式识别这一领域,近两年也引起了国内研究者的注意,已有一些初步的研究工作见诸报道<sup>[2]</sup>。本文将给出我们的一些有意义的结果。

## 2 三元阵列的建立和数据获取

研究表明,阵列的规模和单元的选择是获得好的识别能力的关键之一。但是,作为初步的探索,这里仅建立了一个三元阵列。选用的三种  $\text{SnO}_2$  气敏元件分别为镇江科学仪器厂提供的可燃气体敏感元件,自制的 HM (掺钼)<sup>[3]</sup>型和 JZ<sub>4</sub> (掺钐)型两种元件。阵列的单元电路如图 1 所示。图中  $V_{cc}$  为回路电压,  $GS_i (i = 1 \sim 3)$  为气敏元件单元,  $V_H$  为加热电压,  $R_L$  为负载电阻,  $R_H$  为加热丝电阻。

为了获得甲醇、乙醇、丙酮、乙醚和工业汽油的测试数据,采用了静态测试法。蒸汽浓

1993-08-31 收到, 1994-03-24 定稿

\* 国家自然科学基金资助项目

林海安 男, 1965 年生, 副教授, 现从事固态传感技术、微机械加工、半导体表面、有机无机兼容等领域的研究

吴冲若 男, 1938 年生, 教授, 现从事固态传感技术和功能电子材料(特别是电子陶瓷和超细精密复合功能材料)的研究

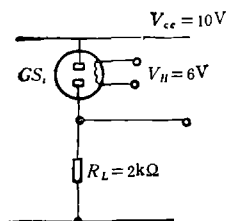


图1 阵列单元电路图

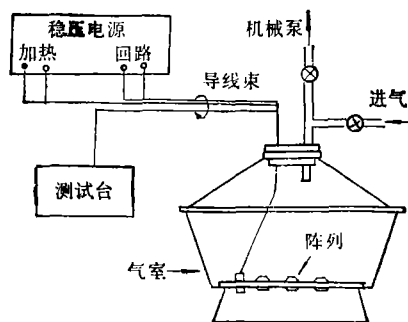


图2 气敏测试系统示意图

度由微量进样器定量,对测试室预抽气,形成压差喷射,从而实现进样和气化。测试系统如图2所示。气敏特性通过  $R_L$  两端电压的改变来反映。为了考虑环境涨落以及元件性能漂移的影响,采集的数据来自4个月范围内的测试,包括了从初春到初夏的季节变化,亦包括了晴、阴、雨三类气候条件。

### 3 模式识别方法

所采用的模式识别框图如图3所示。图左部为学习过程,右部为识别过程。其中预处理阶段主要完成数据整理和必要的转换。

#### 3.1 特征提取

气体的识别是通过比较未知气体与标准气体的特征矢量间的广义距离来完成的,因此模式识别系统的性能强烈地依赖于特征提取的质量。根据  $\text{SnO}_2$  气敏元件的敏感机理<sup>[4]</sup>,有

$$y_{ij} = R_{ij}/R_{0i} = (1 + B_{ij}X_j)^{c_{ij}}, \quad (1)$$

其中  $y_{ij}$  表示阵列中第  $i$  个元件对第  $j$  种气体的响应,  $R_{ij}$  为第  $i$  个元件对第  $j$  种气体响应时的阻值,  $R_{0i}$  为第  $i$  个元件在空气中的阻值,  $X_j$  为第

$j$  种气体的浓度,  $B_{ij}$  和  $c_{ij}$  为与元件和气体种类有关的拟合数据。为了排除浓度项的影响,实现特征提取,(1)式可进一步化为<sup>[5]</sup>

$$y_{ij} = (1 + B_j'X_j)c_{ij}'. \quad (2)$$

将(2)式归一化,有

$$P_{ij} = c_{ij}' / \sum_{i=1}^n c_{ij}' = \log y_{ij} / \sum_{i=1}^n \log y_{ij}, \quad (3)$$

其中  $n$  为阵列单元数,  $j = 1, 2, \dots, m, m$  为气体种类数。与气体浓度无关的气体特征值  $(P_{1j}, P_{2j}, \dots, P_{nj})$  构成了气体特征向量  $P_j$ 。

由于环境和系统的涨落以及元件漂移的影响,对同一气体的  $l$  次测试将得到  $l$  个特

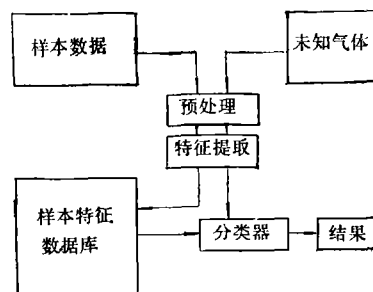


图3 模式识别框图

征向量  $P_i(l)$ , 它们将围绕其平均矢量分布在  $n$  维空间之中。

### 3.2 $k$ -最近邻分类器<sup>[6]</sup>

假设有未知模式  $P$ , 可定义它与  $P_i(l)$  的广义距离为  $D(P, P_i(l))$ <sup>[7]</sup>, 则  $k$ -最近邻分类器的分类规则是根据  $D(P, P_i(l))$  的量度, 如果  $P$  的  $k$  ( $k$  取奇数) 个最近邻模式中的大多数属于  $j$  类, 则判定该未知模式  $P$  属于  $j$  类。通常随着  $k$  值的增大,  $k$ -最近邻分类器的错误概率逼近 Bayes 分类器的错误概率。在本文中, 定义距离  $D$  为欧氏距离。

### 3.3 加权 $k$ -最近邻分类器<sup>[8]</sup>

加权  $k$ -最近邻分类器是对  $k$ -最近邻分类器的改进。它实际上考虑了如果标准样本点越靠近未知模式 (仍然用距离  $D$  量度), 则其对分类结果的影响应该越大。根据上述观点, 可定义权函数

$$\omega_i(l) = [D(P, P_i(l))^2 + 1]^{-1} \times \left\{ \sum_j \sum_l [D(P, P_i(l))^2 + 1]^{-1} \right\}^{-1}, \quad (4)$$

其中  $\omega_i(l)$  满足: (a)  $\sum_{l=1}^k \omega_i(l)$  仅依赖于类  $j$ ; (b)  $\sum_j \sum_l \omega_i(l) = 1$ ; (c)  $\sum_{l=1}^k \omega_j(l) \leq 1$ 。

加权  $k$ -最近邻分类器的分类规律为, 如果

$$\sum_{l=1}^k \omega_j(l) > \sum_{l=1}^k \omega_\lambda(l), \quad (5)$$

其中  $\lambda \neq j, \lambda = 1, 2, \dots, m$ , 则未知模式属于  $j$  类。

## 4 程 序

源程序由 C 语言编写, 在 Turbo C 支持下于微机上调试、编译。

数据库采用了链表结构, 如图 4 所示。该数据库能实现建立、搜索、删除和修改等数据库维护功能。采用这种结构的数据库具有下列特点: (a) 单元数、气体种类、测试次数均可变; (b) 数据结构以单元为基础, 可灵活地实现元件搭配, 完成元件筛选; (c) 只用四个指针变量就可搜索全库。实践表明, 该数据库能较好地满足实际研究中的要求。

进入各功能, 采用了下拉菜单方式, 因此整个程序使用方便。

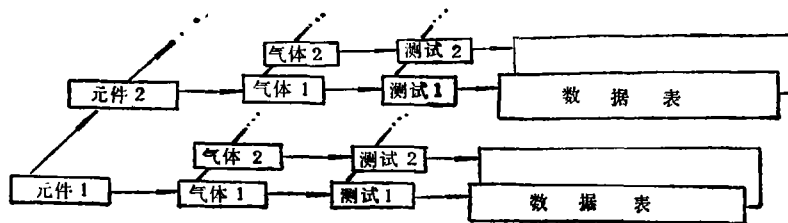


图 4 数据库结构示意图

## 5 结果和讨论

图 5 为甲醇、乙醇、乙醚、丙酮和汽油的特征向量在三个阵列单元上的分布情况。由

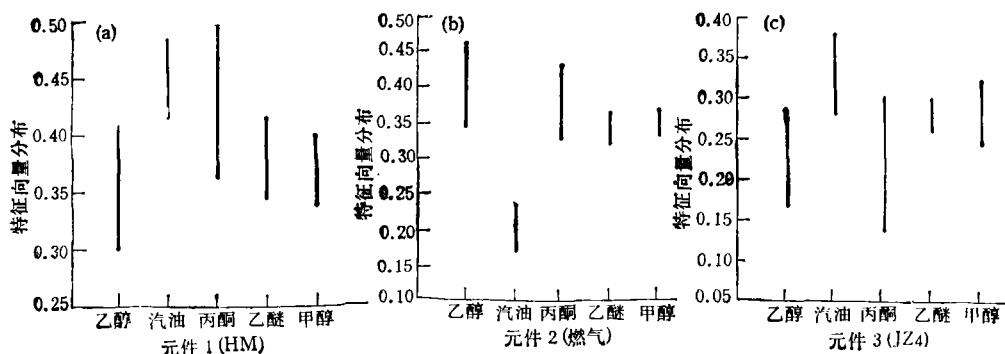


图5 特征向量分布

图可见,大多数气体的特征向量在每一轴上的投影存在不同程度的重叠。但是,从三维空间来看,每种气体有着各不相同的聚类区域,这表明,通过特征提取为模式识别奠定了基础。

表1是2元和3元阵列5-最近邻分类的结果比较,其中2元阵列的结果均不正确,而三元阵列的结果是正确的。可见阵列规模的重要性。通常阵列单元数不应小于需识别的气体种类数<sup>[9]</sup>。研究亦表明,对于特定的使用环境,无限地增加单元数既不可能,也不必要。恰当的阵列规模可以通过实验筛选确定<sup>[10]</sup>。

表1 2元、3元阵列分类比较

元 件	1, 2	2, 3	1, 3	1, 2, 3
结 果	乙 醇	甲 醇	乙 醚	汽 油
实 际	汽 油			

表2 5-最近邻和加权5-最近邻分类器分类结果

气 体	次 数	乙醇		甲醇		乙醚		汽油		丙酮		成功率 (%)	
		A	B	A	B	A	B	A	B	A	B		
乙 醇	10	4	7	4	2	2	1					40	70
汽 油	10					1	1	9	9			90	90

注: A为 $k$ -最近邻, B为加权 $k$ -最近邻。

表2是5-最近邻和加权5-最近邻两种分类结果比较。对乙醇而言,加权5-最近邻分类器的成功率为70%,而5-最近邻分类器的成功率仅为40%,可见加权 $k$ -最近邻分类器具有较好的识别效果。当然,为了进一步提高分类的成功率,必须:(a)增加单元数,至少应大于需识别的气体种类数;(b)采用不同材料制备的气敏元件并对其精心筛选,以便能区分性质相近的气体(如乙醇与甲醇);(c)对阵列单元附加温度调节系统,进一步改善元件工作性能并增强其环境适应性。

## 6 结 束 语

气敏元件阵列和模式识别是改善气敏元件选择性的有效途径。实验表明,加权  $k$ -最近邻分类器具有较高的分类成功率;阵列的规模对分类的成功也有决定性的影响。

### 参 考 文 献

- [1] Persaud K, Dodd G. *Nature*, 1982, 299 (5881): 352—355.
- [2] 梅 涛,何茂彬,采用智能技术提高气敏传感器的综合性能,第二届全国敏感元件与传感器学术会议论文集,上海;1991,870—871.
- [3] 吴冲若,邱洁真,林海安. *电子元件*,1993,35(1): 26—28.
- [4] Clifford P. *Homogeneous Semiconducting Gas Sensors: A Comprehensive Model*. Proc. Meet. on Chemical Sensors, Amsterdam: 1983, 135—146.
- [5] Horner G, Hierold C. *Sensors and Actuators*, 1990, B2(2): 173—184.
- [6] 李介谷,蔡国廉. *计算机模式识别技术*. 上海: 上海交通大学出版社,1986,47—61.
- [7] 张尧庭,方开泰. *多元统计分析引论*. 北京: 科学出版社,1982,393—401.
- [8] 周冠雄. *计算机模式识别. 统计方法*. 武汉: 华中工学院出版社,1986,46—58.
- [9] Shurmer H. *Sensors and Actuators*, 1990, B1(1): 48—53.
- [10] Muller R. *Sensors and Actuators*, 1991, B4(1): 35—39.

## THREE-UNIT GAS-SENSING ARRAY AND ORGANIC SOLVENT DISCRIMINATION

Lin Haian Wu Chongruo Ma Jun

(*Department of Electronic Engineering, Southeast University, Nanjing 210018*)

**Abstract** Three-unit array was constructed with SnO<sub>2</sub> gas sensors. Two classifiers, the  $k$ -nearest neighbour and the weighted  $k$ -nearest neighbour, were used. It is shown that the recognition probability of the former is 40 to ethanol and 90 to gasoline and that of the latter is 70 and 90 respectively. Pattern recognition is an effective method to improve selectivity of gas sensors.

**Key words** Gas sensors, Pattern recognition, Array, Nearest neighbour