

A Method for Gas Qualitative Discrimination Using MEMS Gas Sensor Array Based on Dynamic Classifier Ensemble*

Liu Hang, Tang Zhen'an*

(College of Electronic Science and Technology, Dalian University of Technology, Dalian Liaoning 116024, China)

Abstract: Sensor response drift remains to be the most challenging problem in gas sensing. We proposed a novel ensemble method with dynamic weights to solve the gas discrimination problem regardless of their concentration with high accuracy over extended periods of time. The method uses a dynamic weighted combination of classifiers trained at different points of time. Their weights in testing future datasets are predicted by fitting functions which are obtained by proper fitting of optimal weights in training. We compared the performances of the proposed method and competing methods in experiment based on the public dataset over a period of three years. As results illustrate, the proposed method performs better than others. Furthermore, the method can be further optimized by applying a fitting function that is better match variation of the optimal weight over time.

Key words: MEMS gas sensor; sensor array; drift compensation; classifier ensemble; dynamic weights
EEACC: 7230L **doi:** 10.3969/j.issn.1004-1699.2013.12.005

基于动态分类器集成的 MEMS 气体传感器阵列的气体定性识别方法*

刘 航, 唐祯安*

(大连理工大学电子科学与技术学院, 辽宁 大连 116024)

摘 要: 气体识别研究中, 传感器的性能漂移问题始终是最具挑战的问题之一。为了减少漂移对气体识别的影响, 利用动态分类器集成技术, 针对基于 MEMS 技术的气体传感器阵列, 提出了气体识别方法。主要关注对气体样本在任意浓度下的定性分析。该方法基于支持向量机(SVM)分类器, 首先利用在不同时间段采集的数据分别训练 SVM 分类器, 再利用各分类器对不同时期数据的最优权值, 估计拟合函数的参数; 然后利用拟合函数, 根据数据的采集时刻, 预测各分类器的权值; 最终利用预测的权值, 对所有分类器的识别结果进行集成, 得到最终识别结果。利用累积三年的测量数据, 对该方法与已有类似方法的性能进行了比较。结果显示, 该方法可以在较长时间内具有更高的准确率, 而且, 该方法可以通过选择更合理的拟合函数, 提高识别性能。

关键词: MEMS 气体传感器; 传感器阵列; 漂移补偿; 分类器集成; 动态加权

中图分类号: TP212

文献标识码: A

文章编号: 1004-1699(2013)12-1649-06

电子鼻是一种由气体传感器阵列构成的, 具有模式识别功能的, 气体或气味的检测系统^[1]。针对不同应用领域, 可以定性分析气体成分, 或定量分析各成分浓度。理论上, 气体传感器对相同气体混合物能够得到完全一样的响应信号, 并在测试气体消失后, 响应信号迅速回归基线。然而, 在真实环境下, 目前的传感器在较长时间内无法保持上述性能^[2]。传感器在相同的环境下, 其响应会发生一定的渐变或突变, 也就是所谓的漂移, 这已经被公认为 MEMS 气体传感

器发展中遇到的最严峻的挑战之一^[3-4]。

一般来说, 传感器性能的漂移由两个因素引起^[5]。一个是由传感器的纳米结构决定的, 发生在传感器材料的表面, 被称为一阶漂移(First-Order Drift)。微观结构的变化(如老化或中毒), 影响了传感器表面在与被检测气体接触过程中发生的物理和化学反应。另一个是由测量环境的一些不可控因素产生的漂移, 被称为二阶漂移(Second-Order Drift), 例如温度和湿度的改变、灰尘、各种噪声的叠

项目来源: 国家自然科学基金项目(61131004); 中央高校基本科研业务费专项资金项目(DUT11RC(3)74)

收稿日期: 2013-10-08 **修改日期:** 2013-11-18

加、残留气体的影响等。Romain 等人历经三年对 MEMS 气体传感器阵列进行了测试,得到了非常全面地数据,并系统和深入地分析了传感器性能的漂移现象^[6]。在实际应用中,很难区分由上述两个因素造成的漂移。目前的研究中,一般认为利用可靠的测量系统并在严格的操作控制下,可以忽略二阶漂移的影响,因此对传感器漂移的研究主要是针对一阶漂移。下文中的漂移均指一阶漂移。

为了减小 MEMS 气体传感器的性能漂移,研究者们提出了一些新的传感器材料和纳米结构^[7-8]。这些材料或结构可以在被测气体消失后,更快的脱附气体分子。周期性改变传感器的温度,可以清理传感器表面,有效减少因中毒引起的不可逆吸附^[9]。提高传感器材料的吸附可逆性,可以减弱漂移现象。但是随着时间的延续,漂移的影响不可避免的会逐渐增加。因此,有必要对传感器在较长时间内的特性变化进行研究^[4]。本文关注的是在漂移不可避免的前提下,如何在长时间减小漂移对传感器气体识别的影响,从而延长传感器的使用寿命。

对传感器漂移补偿的最常用方案可以分为单独对每个传感器的补偿和对传感器阵列的补偿。在对传感器阵列的补偿方法中,早期主要利用无监督的成分校正方法(Unsupervised Component Correction),如^[10]。这些方法通过线性变换对传感器的响应序列进行标准化,然后直接利用分类器对结果进行分类。这些方法的主要缺点是它们假设漂移在特征空间是线性的,由此认为通过对数据进行线性变化可以校正漂移。然而,目前已经证明这种假设是不合理的^[11],即性能的漂移是非线性的。因此,这些方法在传感器领域已经很少被使用。另外,该方法中大部分需要利用参考气体计算传感器漂移,并用对参考气体的漂移作为对所有被测气体的漂移。

目前,基于分类器集成的方法在该领域受到越来越多的关注。该类方法首先利用不同数据集训练分类器,得到若干分类器模型。然后,将这些子分类器对待测数据的识别结果进行集成,得到最终结果。文献^[12]已经证明,在漂移补偿方面,分类器集成的性能要优于单个分类器。分类器的集成过程可以分为两类:一类是静态集成,即每个子分类器的权重是固定的;另一类是动态集成,即每个子分类器的权重是随着环境的改变而变化的。

Vergara 等人基于静态集成提出了应对传感器漂移的方法,并且指出分类器集成的方法是对已有成分校正方法的补充,而不是取代^[4]。王海勋等人也提出了类似的分类器集成方法,只是在权重分配方面有所

不同^[12]。这些分类器集成方法利用了静态的权重,相比于单一分类器可以得到更高的准确类。

本文提出了一种基于动态权重的分类器集成方法。实验结果显示,分类器动态集成的准确率随时间的下降速度要慢于静态集成,换句话说,在相同时间段内动态集成的准确率更高。因此,该方法可以在更长时间内缓解漂移对传感器性能的影响,延长传感器的使用寿命。接下来,本文将首先介绍针对纳米气体传感器已提出的分类器集成方法以及关键问题;接下介绍本文提出的动态分类器集成方法;然后通过实验对比和分析了该算法的性能;最后是对本文工作的总结和展望。

1 相关工作

1.1 实验数据

本文中的实验采用了由 Vergara 等人测量并公开的数据集^[4]。传感器阵列由 4 种共 16 个(每种 4 个)气体传感器组成,型号分别为 Figaro 公司的 TGS2600、TGS2602、TGS2610 和 TGS2620。测试系统由 LabVIEW 软件自动控制,避免了人为干扰。测试气体包括氨气、乙醛、丙酮、乙烯、乙醇和甲苯,气体浓度在 5×10^{-6} V ~ $1\ 000 \times 10^{-6}$ V 范围。提取的信号特征包括:

(1) 从待测气体出现到浓度稳定过程中传感器的电阻(或电导)最大变化值,即最大值与基线的差^[13],

$$\Delta R = \max_k r[k] - \min_k r[k] \quad (1)$$

(2) 上述特征的变化率^[14],

$$\|\Delta R\| = \frac{\max_k r[k] - \min_k r[k]}{\max_k r[k]} \quad (2)$$

其中, $r[k]$ 表示传感器电阻的时间序列, k 表示离散的时间索引。

(3) 利用指数滑动平均值(ema-alfa)可以描述传感器的瞬态特征。这是一个经济计量学中的概念,文献^[15]首次将其引入到纳米气体传感器的信号特征提取。其利用 ema-alfa 的最大值(或最小值) $f_\alpha\{r[\cdot]\}$ 来反映序列 $r[\cdot]$ 的增长(或衰减)特征。计算方法如下,

$$y[k] = (1-\alpha)y[k-1] + \alpha(r[k] - r[k-1]) \quad (3)$$

其中, α ($\alpha \in [0, 1]$) 为平滑系数,初始状态 $y[0] = 0$ 。

在测试气体的一次通入/排出过程中,由每个传感器的响应信号提取八个特征,包括 ΔR 、 $\|\Delta R\|$, 以及当 $\alpha = 0.1, 0.01$ 和 0.001 时 ema-alfa($r[k]$) 的最大值和最小值。其中, ΔR 和 $\|\Delta R\|$ 是传感器稳

态响应信号的特征; $\text{ema-alfa}(r[k])$ 的最大值和最小值是瞬态响应信号的特征,反映传感器表面对气体分子的吸附和脱附过程。因此,每次测试可以由传感器阵列获得一个 128 维的特征向量。测试历经三年,共计 13 910 次测量。为了满足分类器的训练集中每个类别有足够且均匀的数据量,已将 36 Mon 的数据分成 10 个数据集。关于测量系统的详细内容请参见文献[4,15]。

1.2 已有分类器集成方法

目前,已经提出多种分类器集成方法^[4,12,14,16],其一般形式如算法 1 所示。对于一个分类问题,输入 x 为特征向量,输出 y 为类标签,在时间段 t 内,得到一个数据集 $S_t = \{X_t, Y_t\} = \{(x_1, y_1), \dots, (x_{m_t}, y_{m_t})\}$,其中 m_t 为特征向量个数。利用 S_t 训练分类器,可以得到 t 时刻的分类器模型 $f_t(x)$ 。对于 $T+1$ 时刻的分类问题,利用先前各分类器的判决结果的加权总和做最后的判决,即 $h_{T+1}(x) = \sum_{i=1}^T \beta_i f_i(x)$,其中 $\{\beta_1, \dots, \beta_T\}$ 为各子分类器的权重。目前,普遍采用的假设是 $T+1$ 时刻的特征分布与前一时刻(即 T 时刻)最接近,因此,已有的分类器集成方法普遍利用数据集 S_T 估计 $\{\beta_1, \dots, \beta_T\}$,只是各自的估计方法不同。

算法 1 分类器集成算法

输入:数据集 $S_t = \{(x_1, y_1), \dots, (x_{m_t}, y_{m_t})\}, t=1, \dots, T$;
 1: for $t=1, \dots, T$ do
 2: 利用数据集 S_t 训练分类器,得到 f_t ;
 3: end for
 4: 利用数据集 S_T 估计 $\{f_1, \dots, f_T\}$ 的权重 $\{\beta_1, \dots, \beta_T\}$;
 5: 对 $\{\beta_1, \dots, \beta_T\}$ 归一化;
 输出:子分类器模型 $\{f_1, \dots, f_T\}$ 和对应的权重 $\{\beta_1, \dots, \beta_T\}$ 。

分配权重的方法有很多。一个简单方法是依据每个分类器对 S_T 的分类准确率,如文献[4]。王海勋等人将 $\beta_i = \text{MSE}_i - \text{MSE}_i$ 作为分类器 f_i 的权重,其中, MSE_i 是 f_i 对 S_T 的均方误差, MSE_i 是分类器对 S_T 随机预测的均方误差^[12]。在确定每个分类器的权重后,根据对各分类器的判决结果进行加权求和,做出最后判决。

然而,实验结果显示已有方法得到的权值并不是最优的。实验设置如下,在 T 时刻,利用数据集 S_1, \dots, S_T 分别训练 SVM 分类器,得到分类器模型 f_1, \dots, f_T 。然后利用最优权重和由上述方法得到的权重分别组成分类器集合,对 S_T 的数据进行分类。最优权重是通过遍历搜索所有权重组合,得出的可

以使分类准确率最大的权重。采用最优权重的分类器集成的性能,是这些子分类器利用集成方法在理论上能够得到的最优性能。图 1 显示了当 $T=2, \dots, 10$ 时,利用各种权重得到的分类准确率。从图中可以看出,基于已有权重估计方法的分类性能与最优性能有很大差距。

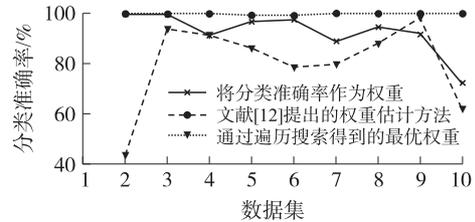


图 1 利用不同权重分配方法的分类器集成的性能

基于静态加权的分类器集成方法,各分类器的权重在确定后不再改变。也就是说,利用 S_T 估计的权重被用于对所有后续数据的检测。对于数据集 S_{T+n} 而言,如果 n 较小,根据气体传感器的原理, S_{T+n} 与 S_T 的数据分布会比较接近,分类器对 S_{T+n} 的分类准确率仍然可以保持在较高水平。但是,随着 n 的增加,传感器在 T 和 $T+n$ 时刻的相对漂移范围在增加,对分类器的影响会逐渐增强,结果就是分类准确率随着时间跨度的增加而不断降低。

为了验证上文中提到的相对漂移对分类性能的影响,我们进行了下面的实验。实验中,将 $S_1 \sim S_5$ 作为 5 个训练集, $S_6 \sim S_{10}$ 作为测试集。首先,利用训练集得到 5 个 SVM 分类器模型,即 $f_1 \sim f_5$ 。然后利用文献[4,12]中的方法(记为方法 1 和 2),以及均匀加权方法(方法 3)将分类器集成。为了衡量这些方法的性能,我们通过遍历搜索得到了 $f_1 \sim f_5$ 的集成对每个测试集在理论上可以达到的最优性能(方法 5)。受遍历搜索的运行时间限制,每个权重的搜索范围设为 $[0, 1]$,搜索间隔设为 0.01。另外,为了验证上文提到的假设,我们用每个数据集训练分类器后,用得到的单一 SVM 模型分类下一个数据集(方法 4)。所有方法对测试集的分类性能如图 2 所示,其中横坐标表示收集每个数据集的平均时刻。从图中可以看出,方法 4 的性能在大部分时间都不是最

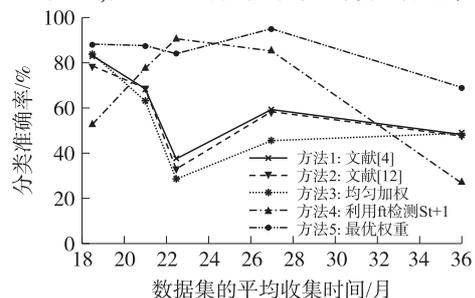


图 2 不同方法的分类性能

好的。这说明相比于由近期数据训练得到的单一分类器,更早期的分类器集成如果权重适当,可以得到更好的性能。

在实际应用中估计子分类器在未来不同时刻的最优权重是非常困难的,本文的研究目标是预测子分类器在未来某时刻的次优权重。方法 1 ~ 方法 3 的性能在初始阶段比较接近最优性能,但是随着时间的推移,性能在逐渐下降。这说明短时间内的次优权重可以根据当前数据估计得到,但是随着训练集和测试集之间时间跨度的增加,其性能会不断降低。因此,本文为了减缓分类器集成的性能的下降速度,提出了基于动态分类器集成的气体传感器阵列信号的漂移补偿算法。

2 动态分类器集成算法

使用静态权重的分类器集成的性能,会随着训练集与测试集的采集时间间隔的增加而降低。本节介绍一种新的基于动态加权的分类器集成算法,目的是减缓分类性能下降的速度,该算法的一般形式如算法 2 所示。

在训练阶段,由每个时间段 $t(t \leq T)$ 内的数据集 S_t 均可以训练得到一个子分类器模型 f_i 。其中, T 为当前时刻索引,训练集 $S_t = \{X_t, Y_t\} (t \leq T)$ 中的所有数据已知。因此,在训练阶段,对于任一数据集 S_t ,都可以用全部子分类器检测,无论这个子分类器是由 S_t 之前的数据集训练得到,还是之后的得到。另外, w_t 为时间段 t 内数据集 S_t 的平均采集时刻,在拟合阶段将作为拟合函数的输入,利用不同拟合函数针对每个子分类器 f_i ,建立数据采集时刻与该子分类器的最优权重 β_i^t 间的数学关系,即算法 2 中训练与拟合阶段的第(9)步。拟合函数用 $C_i(w)$ 表示,其中 t 为对应的子分类器索引,即 f_i, w 为拟合函数的输入。从而在后续的检测阶段($t > T$),将数据采集时刻作为输入,利用拟合函数用 $C_i(w)$ 估计每个子分类器的最优权重。在检测阶段,所有结果都是通过集成这些子分类器的结果而得到。

算法 2 动态分类器集成算法

(1) 训练与拟合阶段

输入:训练集 $S_t = \{(x_1, y_1), \dots, (x_{m_t}, y_{m_t})\}, t = 1, \dots, T$, 以及索引 t 对应的每个数据集平均采集时刻 w_t ;

```

1 for  $t = 1, \dots, T$  do
2   利用数据集  $S_t$  训练分类器模型,得到  $f_i$ ;
3 end for
4 for  $t = 1, \dots, T$  do
5   估计分类器组  $\{f_1, \dots, f_T\}$  对于数据集  $S_t$  的最优权重集

```

合 $\{\beta_1^t, \dots, \beta_T^t\}$;

```

6 end for
7 得到一个  $T \times T$  的权重矩阵  $\beta$ , 其第  $i$  行第  $j$  列的元素  $\beta_i^j$  表示子分类器  $f_i$  对数据集  $S_j$  的最优权重;
8 for  $t = 1, \dots, T$  do
9   利用二维数组  $[(w_1, \beta_1^t), \dots, (w_T, \beta_T^t)]$  估计拟合函数  $C_i(w)$  的参数;
10end for
输出:分类器模型  $f_i$  和对应的拟合函数  $C_i(w), t = 1, \dots, T$ 。

```

(2) 检测阶段

输入:数据集 X_{T+n} 及其平均采集时刻 $w_{T+n}, n > 0$; 分类器模型 f_i 和对应的拟合函数 $C_i(w), t = 1, \dots, T$;

```

1 for  $t = 1, \dots, T$  do
2   利用拟合函数  $C_i(w)$  计算分类器  $f_i$  在  $w_{T+n}$  时刻的最优权重,即  $C_i(w_{T+n})$ ;
3 end for
4 利用分类器集成  $h_{T+n} = \sum_{i=1}^T C_i(w_{T+n}) f_i$  检测  $X_{T+n}$ ;
输出:估计得到的类标签  $Y_{T+n} = h_{T+n}(X_{T+n})$ 。

```

整个算法存在两个关键问题。一个是在训练阶段如何获得所有子分类器对每个数据集的最优权重集合。实验结果(图 1)已经证明,使用已有算法估计得到的权重所达到的分类性能,与理论上的最优性能差距较大。本文通过遍历搜索,得到所有子分类器 $f_i (i = 1, \dots, T)$ 对每个训练集 $S_j (j \leq T)$ 的最优权重集合。遍历搜索可以确保权重最优,但是运行时间随 T 的增大呈指数增加。由于遍历搜索运行时间的限制,对每个权重的搜索范围设定在 $[0, 1]$, 搜索间隔为 0.05, 且训练集不能过多。表 1 显示了当 $T = 5$ 时,所有子分类器对不同时刻数据集的最优权重。每个子分类器 f_i 对数据集 S_j 的最优权重,构成了最优权重矩阵 β 。

表 1 当 $T = 5$ 时,所有子分类器对不同时刻数据集的最优权重,即 β_i^j

子分类器, 即 f_i	数据集索引(数据集的平均采集时刻),即 $j(w_j)$				
	1(1.5)	2(6.5)	3(12)	4(14.5)	5(16)
f_1	0.45	0.05	0.15	0.05	0.05
f_2	0.05	0.45	0	0	0.15
f_3	0.10	0	0.75	0.25	0.15
f_4	0.25	0.10	0.10	0.65	0
f_5	0.15	0.40	0	0.05	0.65

从 β_i^j 可以发现,每个子分类器的最优权重是随时间变化的。子分类器 f_i 在时间段 j 的最优权重和 j 对应的真实时刻组成了一个二维数组 (β_i^j, w_j) , 该

数组在训练阶段 $0 < j \leq T$ 的变化,即 $\{(\beta_i^1, w_1), \dots, (\beta_i^T, w_T)\}$,可以由一个关于时间 w 的函数 $C_i(w)$ 拟合得到。在检测阶段,子分类器 f_i 在 w_{T+n} ($n=1, 2, \dots$) 时刻的最优权重由 $C_i(w)|_{w=w_{T+n}}$ 预测得到。

本算法的第二个关键问题是如何构建拟合函数。理论上, f_i 由 i 时刻的数据训练得到,因此 f_i 最符合该时刻的数据特征,那么 f_i 在 i 时刻的权重应该最大;另外, i 时刻两侧的数据集,随着与 i 时刻数据的时间间隔的增加,其数据特征受传感器信号漂移的影响逐渐增大, f_i 对其的分类性能理论上应该逐渐降低,因此权重应该逐渐减小。因此,本文选择的拟合函数应该满足下列条件:(1) $C_i(w)$ 在 w_i 时刻具有最大值;(2) $C_i(w)$ 向 w_i 两侧递减;(3) $C_i(w) \geq 0$ 。根据上述条件,本文选择的拟合函数为:

$$y = \frac{c}{\exp(ax+b) + \exp(-(ax+b))} + d$$

通过训练阶段的拟合过程,估计得到上述函数的 4 个参数,即 a, b, c 和 d 。

在测试阶段,对于在 w_{T+n} 时刻的特征向量,每个子分类器 f_i 的权重为 $C_i(w_{T+n}), i=1, \dots, T$ 。最终的检测结果为各子分类器结果的集成:

$$y_{T+n} = \sum_{i=1}^T C_i(w_{T+n})f_i(x_{T+n})$$

3 实验结果

本文利用 SVM 分类器及公开数据进行了方法验证及性能对比。在公开的数据中,由于甲苯的数据缺失近一年,因此本文利用另外 5 种气体的数据对所提算法进行验证。与上文中的实验类似,本节仍选择数据集 $S_1 \sim S_5$ 作为训练集, $S_6 \sim S_{10}$ 作为测试集。

依据所提算法的训练与拟合过程,得到 5 个子分类器 $f_1 \sim f_5$ 以及对应的拟合函数 $C_1(w) \sim C_5(w)$,拟合函数的参数如表 2 所示。由拟合函数估计得到的,各子分类器在测试集数据的采集时刻的权重如表 3 所示。图 3 显示了本文算法的分类准确率(方法 7),与之对比的分类器集成方法包括

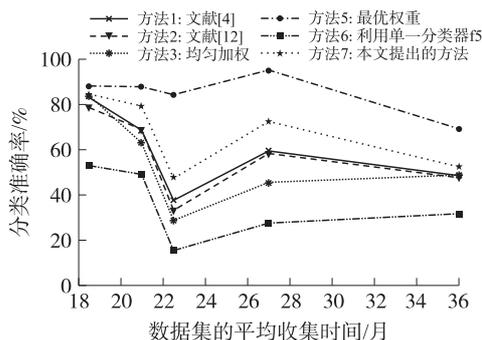


图 3 各种分类方法的性能对比

图 2 中的方法 1 ~ 方法 3。由于训练集 S_5 在采集时间上与测试集最接近,因此图 3 也显示了利用 f_5 作为单一分类器的分类准确率(方法 6)。另外,图 3 还显示了通过遍历搜索得到的, $f_1 \sim f_5$ 的集合对 $S_6 \sim S_{10}$ 在理论上能得到的最高准确率(方法 5)。

从实验结果可以看出,分类器集成的性能均高于单一分类器。在集成方法中,方法 1 和方法 2 的性能相当,且略高于方法 3,但是都明显低于本文所提方法。因此,本文所提方法相比于这些方法,可以更进一步的延长气体传感器的使用寿命。

表 2 5 个拟合函数的参数

子分类器, 即 f_i	拟合函数的参数			
	a	b	c	d
f_1	797	0.155	7.816	-0.013 14
f_2	-11.01	1.252	3.062	0.211 5
f_3	-15.99	1.877	-7.713	0.422 5
f_4	-0.281 6	0.154 6	-0.845 8	0.346 6
f_5	0.497 1	2.576	-7.546	0.175

表 3 各拟合函数预测的每个子分类器的权重

子分类器, 即 f_i	数据集索引(数据集的平均采集时刻),即 $j(w_j)$				
	6(18.5)	7(21)	8(22.5)	9(27)	10(36)
f_1	0.021 2	0.010 4	0.005 8	-0.003 0	-0.009 4
f_2	0.191 4	0.189 4	0.188 5	0.186 8	0.185 6
f_3	0.382 3	0.378 4	0.376 6	0.373 2	0.370 6
f_4	0.246 8	0.265 1	0.273 1	0.288 4	0.299 7
f_5	0.158 3	0.156 7	0.156 0	0.154 6	0.153 5

对于本文所提算法,训练集的个数会影响拟合函数的参数估计,从而通过影响子分类器的权重估计,最终影响分类器集成的性能。为了验证上述观点,我们进行了下面的实验。首先,我们分别利用 4, 5, 6, 7 和 8 个数据集(分别为 $S_5 \sim S_8, S_4 \sim S_8, S_3 \sim S_8, S_2 \sim S_8$ 和 $S_1 \sim S_8$) 作为训练集,构建集成分类器后对 S_9 和 S_{10} 中的数据进行分类,每个集成传感器的准确率如图 4 中曲线所示。图 4 中的直方图显示了每个集成传感器使用的训练集的时间跨度。从结果可以看出,分类准确率随训练集数量的增加有所

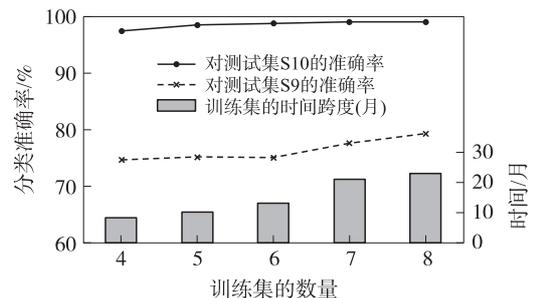


图 4 集成不同数量子分类器的分类准确率

增加,而且训练集与测试集的时间间隔越大,准确率增加越明显。因此,本文算法可以通过增加训练集的数量(每个数据集的时间跨度不变),实现提高检测准确率。换句话说,增加训练数据的采集时间跨度,可以提高气体传感器在检测阶段的使用时间。

4 结论

为了针对 MEMS 气体传感器阵列,减小漂移对气体识别性能的影响,本文提出了新的动态分类器集成的方法。实验结果显示,本文的方法相比于已有方法,可以在更长时间内获得较高分类准确率。该算法有两个关键问题。一个是在训练阶段,如何估计每个子分类器的最优权重。实验显示,目前的方法不能获得最优权重,而且与最优性能的差距较大。遍历搜索可以获得最优权重,但是运行速度限制了其实际应用。另一个关键问题是构建更合适的拟合函数。我们将在接下来的工作中,进一步研究上述问题。

另外,我们正利用与文献[4,15]介绍的相似的系统采集气体传感器对5种挥发性气体的响应信号,并且记录了每次采样的时间、温度和湿度信息。目的是在拟合函数中增加温度和湿度信息,提高对最优权重的预测性能。并在测试阶段,可以得到每个样本的采集时刻,使得对子分类器权重的预测更精确。

参考文献:

- [1] Gardner J W, Bartlett P N. *Electronic Noses: Principles and Applications* [M]. New York: Oxford University Press, 1999.
- [2] Holmberg M, Winquist F, Lundström I, et al. Drift Counteraction for an Electronic Nose [J]. *Sensors and Actuators B: Chemical*, 1996, 36(1): 528-535.
- [3] Zuppa M, Distanto C, Siciliano B, et al. Drift Counteraction with Multiple Self-Organising Maps for an Electronic Nose [J]. *Sensors and Actuators B: Chemical*, 2004, 98(2): 305-317.
- [4] Vergara A, Vembu S, Ayhan T, et al. Chemical Gas Sensor Drift Compensation Using Classifier Ensembles [J]. *Sensors and Actuators B: Chemical*, 2012, 166: 320-329.

- [5] Holmberg M, Artursson T. Drift Compensation, Standards, and Calibration Methods [M]. *Handbook of Machine Olfaction: Electronic Nose Technology*, 2004: 325-346.
- [6] Romain A, Nicolas J. Long Term Stability of Metal Oxide-Based Gas Sensors for E-Nose Environmental Applications: An Overview [J]. *Sensors and Actuators B: Chemical*, 2010, 146(2): 502-506.
- [7] Pandey N K, Tiwari K, Roy A, et al. Ag-Loaded WO₃ Ceramic Nanomaterials: Characterization and Moisture Sensing Studies [J]. *International Journal of Applied Ceramic Technology*, 2013, 10(1): 150-159.
- [8] Kim I, Rothschild A, Tuller H L. Advances and New Directions in Gas-Sensing Devices [J]. *Acta Materialia*, 2013, 61(3): 974-1000.
- [9] Gosangi R, Gutierrez-Osuna R. Active Temperature Modulation of Metal-Oxide Sensors for Quantitative Analysis of Gas Mixtures [J]. *Sensors and Actuators B: Chemical*, 2013, 185(0): 201-210.
- [10] Artursson T, Eklöv T, Lundström I, et al. Drift Correction for Gas Sensors Using Multivariate Methods [J]. *Journal of Chemometrics*, 2000, 14(5-6): 711-723.
- [11] Schölkopf B, Smola A, Müller K. Nonlinear Component Analysis as a Kernel Eigenvalue Problem [J]. *Neural Computation*, 1998, 10(5): 1299-1319.
- [12] Wang H, Fan W, Yu P S, et al. Mining Concept-Drifting Data Streams Using Ensemble Classifiers [C] // *Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Washington, DC, USA, 2003: 226-235.
- [13] Llobet E, Brezmes J, Vilanova X, et al. Qualitative and Quantitative Analysis of Volatile Organic Compounds Using Transient and Steady-State Responses of a Thick-Film Tin Oxide Gas Sensor Array [J]. *Sensors and Actuators B: Chemical*, 1997, 41(1): 13-21.
- [14] Vergara A, Huerta R, Ayhan T, et al. Gas Sensor Drift Mitigation Using Classifier Ensembles [C] // *Proceedings of the Fifth International Workshop on Knowledge Discovery from Sensor Data*, New York, NY, USA, 2011: 16-24.
- [15] Muezzinoglu M K, Vergara A, Huerta R, et al. Acceleration of Chemosensory Information Processing Using Transient Features [J]. *Sensors and Actuators B: Chemical*, 2009, 137(2): 507-512.
- [16] Kadri C, Tian F, Zhang L, et al. Neural Network Ensembles for Online Gas Concentration Estimation Using an Electronic Nose [J]. *International Journal of Computer Science Issues*, 2013, 10(2): 129-135.



刘航(1982-),男,博士,讲师,主要研究领域为半导体传感器及其应用, liuhang@dlut.edu.cn;



唐祯安(1955-),男,博士,教授,博士生导师,主要研究领域为半导体传感器及其应用、集成电路与系统, tangza@dlut.edu.cn。