

文章编号:1000-582X(2013)04-075-05

用于气敏传感器漂移补偿的数据选择方法

刘 涛,黄智勇

(重庆大学 通信工程学院,重庆 400044)

摘 要:针对传感器在线漂移补偿中可能存在多种状态的样本,进而引起误判的情况,提出了一种新的数据剔除与判别方法,该方法将传感器阵列响应曲线斜率作为依据进行数据剔除,同时将输入样本与模式识别算法中的记忆知识进行对比,从而进行数据归类,以此避免识别错误的发生。实验表明:该方法配合在线漂移补偿方法后,能够自动判断当前样本所处状态。识别正确率由 37.5% 提升至 100%。

关键词:数据处理;化学传感器;漂移;数据剔除

中图分类号:TP212

文献标志码:A

A data selection method applied to on-line drift compensation for gas sensors

LIU Tao, HUANG Zhiyong

(College of Communication Engineering, Chongqing University, Chongqing 400044, China)

Abstract: A new data culling and labeling method is proposed to avoid misleading outcomes caused by multi-state samples during drift compensation process. This method culls data by the curve slopes of gas sensor array response and labels data by comparing input samples with the memories of pattern recognition algorithm to avoid occurrence of misleading results. Experiments show the method combined with on-line drift compensation algorithm can estimate sample-states automatically and increase the recognizing accuracy from 37.5% to 100%.

Key words: data processing; chemical sensors; drift; data culling

随着社会的发展,气敏传感器作为人工嗅觉系统的重要组成部分受到了广泛关注^[1]。相比传统的气体检测方法,由气敏传感器阵列和模式识别算法组成的电子鼻系统具有检测速度快,操作简便等优势,在医疗卫生^[2],环境保护^[3],食品检测^[4]等领域拥有巨大的应用潜力。

电子鼻设备工作方式可大致分为离线^[5]和在线检测^[6]2 大类:离线检测可以将完整的检测数据进行收集和整理,然后利用相应的数学方法进行识别;而在线检测是一系列连续的实时判断过程,每一次判断只能依靠过去时刻的有限数据且其判断结果可以作

为后续时刻的判断依据。对于需要进行长期不间断检测的对象,常采用在线检测方式,在这种方式下,随着工作时间的延长,由于老化、中毒等因素,电子鼻的气敏传感器阵列将会产生长期漂移^[7]。长期漂移的存在使得电子鼻系统的识别精度受到影响。

为了克服长期漂移带来的问题,相关研究也陆续展开。有学者提出使用各种统计方法将漂移引起的变化从原始信号中剔除,从而获得正确的信号^[8-10],然而这种方法往往需要大量检测数据作为先验信息^[11],因此更适合于离线检测。对于在线检测中的漂移补偿问题,通常使用自适应补偿的方法

收稿日期:2012-11-22

基金项目:重庆市自然科学基金资助项目(CSTC2012JJA4005)

作者简介:刘涛(1980-),男,重庆大学博士研究生,主要研究方向智能信号处理、人工神经网络、人工嗅觉技术;

(Tel) 13983422945; (E-mail) cquliutao@cqu.edu.cn。

加以解决,如:径向基神经网络^[12]、自组织映射图(self organizing map, SOM)神经网络^[13-14]等,其中 SOM 网络由于其结构简单,计算量小,因而应用较广,但 SOM 神经网络的单层结构无法避免各类样本在训练中对各自区域的相互影响,所以,近年来,一种由多个 SOM 网络组成的多重自组织映射图(multiple self organizing map, MSOM)网络^[15]逐渐开始应用于电子鼻漂移在线补偿的研究中并取得了满意效果^[16-17]。

针对电子鼻在线漂移补偿问题,首先通过理论分析指出常规的 MSOM 方法由于不能辨识数据状态而存在盲目重训练现象。由此提出了一种数据选择方法,该方法通过提取响应曲线形态学特征和模式识别网络中的记忆信息,进行样本筛选,避免错误识别结果的产生。最后通过对比实验证明了新方法在在线漂移抑制方面的有效性和先进性。

1 原理与方法

1.1 基于多重自组织图的在线漂移补偿方法

MSOM 由多个 SOM 平面组成,平面数量与样本类别数对应相等。训练时,各种样本单独训练对应的 SOM 平面,其训练方法与 SOM 网络训练方法一致。

为了能够在漂移补偿中使用 MSOM 网络,因此在网络的判别过程中加入了重训练过程,其步骤如下

1)按照下式计算 t 时刻输入样本 $\mathbf{X}(t)$ 对应的网络输出

$$\begin{cases} O(t) = f(\min \|\mathbf{X}(t) - \mathbf{w}_{km}(t)\|), \\ O(t)' = f(\min \|\mathbf{X}(t) - \mathbf{w}_{km}(t)\|'), \end{cases} \quad (1)$$

其中: \mathbf{w}_{km} 代表第 k 个 SOM 网络中第 m 个神经元权值, $\min \|\mathbf{X}(t) - \mathbf{w}_{km}\|$ 表示按照距离最短原则得到的获胜神经元, $f(\cdot)$ 用于计算神经元所在平面的编号, $O(t)$ 为获胜神经元所在的 SOM 平面编号,也即是当前 MSOM 的输出; $\min \|\mathbf{X}(t) - \mathbf{w}_{km}(t)\|'$ 表示求距离大小仅次于获胜神经元的次获胜神经元; $O(t)'$ 即是次获胜神经元所在 SOM 平面的编号。

2)若 $O(t) = O(t)' = n$, 那么按照下式进行权值迭代

$$\mathbf{w}_{km}(t+1) = \mathbf{w}_{km}(t) + a \cdot [\mathbf{X}(t) - \mathbf{w}_{km}(t)] \quad \forall k = n, \\ c_{km} \in \varphi, \quad (2)$$

其中 a 为比例系数, c_{km} 表示参与训练的神经元, φ 表示邻域范围,该邻域以获胜神经元为中心并以指数规律减小。

3)在极少数情况下,存在 $O(t) = n, O(t)' = n' \neq$

n , 此时按照下式进行权值迭代

$$\begin{cases} \mathbf{w}_{km}(t+1) = \mathbf{w}_{km}(t) + a \cdot [\mathbf{X}(t) - \mathbf{w}_{km}(t)], \\ \forall k = n, c_{km} \in \varphi, \\ \mathbf{w}_{km}(t+1) = \mathbf{w}_{km}(t) - a \cdot [\mathbf{X}(t) - \mathbf{w}_{km}(t)], \\ \forall k = n', c_{km} \in \varphi', \end{cases} \quad (3)$$

其中 φ' 为以次获胜神经元为中心并以指数规律减小的邻域。

4)令 $t = t + 1$, 返回步骤(1), 开始下一次重训练过程。

1.2 多重自组织图方法的盲目重训练问题

如图 1 所示为气体传感器一次完整检测过程的响应曲线,该曲线为连续函数且其状态可以分为稳态和暂态 2 类,其中暂态过程为 2 类气体样本检测结果之间的过渡状态。显然,MSOM 在漂移补偿中需要用稳态样本作为重训练样本。若不对数据样本加以筛选,那么暂态样本和其他干扰气体样本同样可以进入重训练过程,进而扰乱漂移补偿算法。

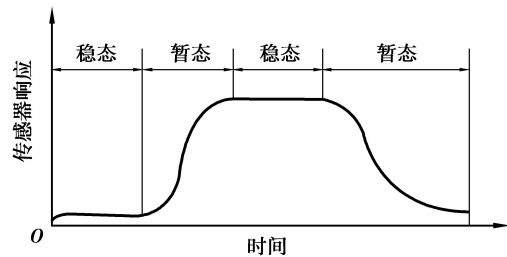


图 1 传感器典型响应曲线

设 t 时刻传感器阵列输出为 $\mathbf{X}(t) = (x_1(t), x_2(t), \dots, x_i(t))$, 其中 $x_i(t)$ 为单个传感器的响应。假设在 $[t_1, (t_1 + b \cdot \Delta t)]$ 期间为第一类气体样本与第二类气体样本间的暂态过程, 即有

$$\begin{cases} O(t) = 1, t < t_1, \\ O(t) = 2, t > (t_1 + b \cdot \Delta t). \end{cases} \quad (4)$$

其中 b 为自然数, Δt 为采样间隔。根据前述 MSOM 算法原理和暂态响应曲线为连续函数的情况, 那么只要采样间隔足够小, 即有

$$\begin{cases} O(t_1 + \Delta t) = 1; \\ O(t_1 + 2 \cdot \Delta t) = 1; \\ \dots \dots \\ O(t_1 + b \cdot \Delta t) = 1; \\ O(t) = 1, t > (t_1 + b \cdot \Delta t); \end{cases} \quad (5)$$

可见,按照 MSOM 算法,式(5)中最后一项将与式(4)第二项矛盾。其原因是没有对暂态过程样本加以区分,造成盲目重训练,以致 MSOM 将正确的记忆信息遗忘。

此外,考虑到在线检测使用环境的不可预知性,当出现未在初始训练过程中出现的干扰气体样本时,由于 MSOM 网络不具备增量学习能力,那么当重训练过程进行后,干扰气体信息将进入网络并对原有记忆信息进行清洗,最终也会导致误判。

综上,如果没有一种有效的数据筛选方法,那么 MSOM 于在线漂移补偿过程中将可能因盲目重训练问题而发生识别错误。

1.3 基于形态学特征和模式知识的数据选择方法

解决盲目重训练问题的根本在于剔除掉对模式识别造成误导的样本,这些样本包括 2 类:暂态样本和干扰气体样本。

笔者提出了一种具有拒绝机制的数据选择方法。该方法配合在线漂移补偿和识别算法,以响应曲线形态学特征——斜率以及后续识别算法中权值所包含的记忆知识为依据,实时判断输入样本状态,从而克服盲目重训练现象的发生。其主要步骤如下(后续漂移补偿和识别算法为 MSOM)

1) 设置响应曲线斜率筛选阈值 θ_φ 和距离筛选阈值 θ_d

$$\theta_\varphi = (1 + c_\varphi) \cdot \text{minslop}_{\text{train}},$$

$$\theta_d = (1 + c_d) \cdot \max \| \mathbf{X}_{\text{train}} - \mathbf{w}_{\text{win_train}} \|,$$

其中 $\text{minslop}_{\text{train}}$ 表示训练样本暂态过程中传感器阵列响应曲线的最小斜率 $\max \| \mathbf{X}_{\text{train}} - \mathbf{w}_{\text{win_train}} \|$ 为训练样本的稳态过程中,样本与获胜神经元的最大欧式距离,考虑到实际检测中存在漂移和抖动现象,引入了调整系数 c_φ 和 c_d ;

2) 输入判别过程样本 $\mathbf{X}(t)$, 根据下式计算输入样本即时斜率 $\text{slop}(t)$

$$\text{slop}(t) = \begin{cases} \| \mathbf{X}(t) - \mathbf{X}(t-n) \|, & t > n \\ \| \mathbf{X}(t) - \mathbf{X}(1) \|, & t \leq n \end{cases},$$

其中 n 为自然数,起平滑斜率的作用,减小信号干扰等因素对斜率的影响;

3) 归一化 $\mathbf{X}(t)$ 并按照下式计算获胜神经元 $\mathbf{w}_{\text{win}}(t)$ 与 $\mathbf{X}(t)$ 的欧式距离 $\Delta d(t)$

$$\Delta d(t) = \| \mathbf{X}(t) - \mathbf{w}_{\text{win}}(t) \|;$$

4) 如果 $\text{slop}(t) > \theta_\varphi$, 那么判断该样本为暂态样本,至步骤(7);

5) 如果 $\Delta d(t) > \theta_d$, 那么判断该样本为干扰气体样本,至步骤(7);

6) 若输入样本满足 $\text{slop}(t) \leq \theta_\varphi$ 且 $\Delta d(t) \leq \theta_d$, 则判断该输入样本为稳态数据,根据式(2)(3)进行重训练;

7) 令 $t=t+1$,至步骤(2),直到停止工作。

通过上述方法,可以根据 θ_φ 判断出暂态过程样

本,根据 θ_d 判断出干扰气体样本,以此避免后续算法受到影响,解决盲目重训练问题。

2 结果与讨论

2.1 实验条件

实验平台如图 2 所示。实验中将被测气体的标气配制在气体采集袋中,然后采用泵吸方式通过泵 1 将被测气体打入测试腔中。传感器阵列由 4 个金属氧化物气体传感器组成,分别为:GSBT11、TGS2620、TGS2602 和 TGS2201,其中 TGS2201 内部集成了 2 个气体传感器,故共有 5 路气体传感器输出。数据采集模块采用 12 位串行 AD 芯片 TLC2543,每次采集时间间隔设置约为 1 s。采集到的数据通过串口传输到 PC 机中保存并进行处理。在实验中可利用泵 2 排出测试腔中的气体并用阀门进行控制。数据处理软件采用 MATLAB7.1。

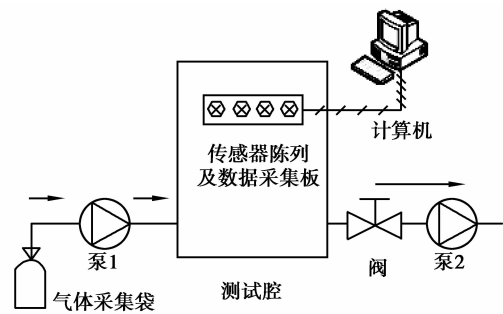


图 2 实验平台结构图

按照时间先后顺序一共选择了 15 次实验的数据,每次实验时间约为 4 min,每次实验采集数据 240 组,共采集数据 3 600 组。每次实验都分为基线、进样和气体稳态 3 个阶段。按时间顺序,进行了 5 次一氧化碳(CO),5 次甲醛(CHO)和 5 次二氧化氮(NO₂)实验。在 MSOM 网络训练时,将 CO(包括基线和气体样本)和 CHO(不包括基线)在第一次实验中的稳态数据作为训练数据,NO₂ 作为干扰气体不参与训练,故 MSOM 网络由 3 个 SOM 平面构成,3 个平面分别对应基线(洁净空气)、CO 和 CHO。参与测试(重训练)的样本为剩下的 13 次实验数据,每次 240 组,共 3 120 组。

2.2 数据预处理

为便于分析,数据预处理分别采用 2 种方式:对同一输入样本 $\mathbf{X}(t)$,求响应曲线的斜率时

$$\mathbf{X}(t) = \frac{\mathbf{X}(t)}{X_{\text{max}}},$$

考虑到 AD 的采样精度为 12 位,故 X_{max} 取值为 4 096。采用这种方式的目的是在不改变单个传感

器响应变化趋势的情况下,将斜率大小控制在一定范围内;求 $\Delta d(t)$ 时

$$\mathbf{X}(t) = \frac{\mathbf{X}(t)}{\|\mathbf{X}(t)\|},$$

其目的是使得样本模值与 MSOM 记忆权值模值相等,以便于进行基于欧式距离的判断。

2.3 数据选择结果

对于前 8 次实验(CO 和 CHO 实验):每次实验,判决层正确输出结果顺序应当为:1(基线)→2(气体进样)→1(气体稳态)→2(排气);对于第 9~13 次实验(NO_2 实验):输出结果的顺序应当为:1(基线)→2(气体进样)→3(干扰气体)→2(排气)。

若用 F 表示数据选择结果。当 $F=1$ 时,代表输入样本为基线、CO 或者 CHO 样本,应当保留;当 $F=2$ 时,代表暂态过程,应当丢弃;当 $F=3$ 时,代表干扰气体样本(NO_2),应当丢弃。

由图 3 可知,当 $\theta_\varphi=0.015, \theta_d=0.1$ 时,数据选择算法判断结果与实际结果基本吻合。在总共 3 120 个判决结果中,共有 6 处判决结果出现波动(见表 1)。其原因为,这 6 处波动的数据样本均处于暂态与稳态过程的交界处,由于传感器数据抖动造成了判断结果的抖动。实验结果表明:采用的数据选择方法是正确而有效的,可以达到剔除暂态过程和干扰气体的目的。

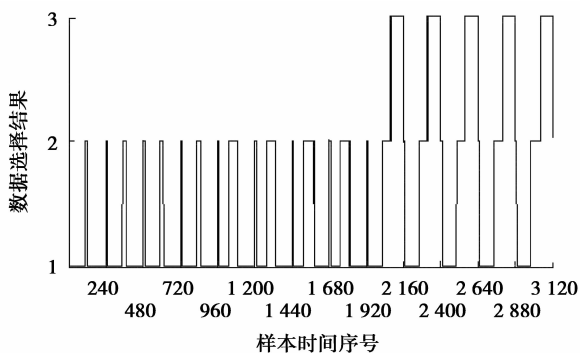


图 3 数据选择结果

表 1 预判决结果中的波动

波动样本 时间序号	当前样本输出	前后相邻 样本输出
第 1088 号	1	2
第 1333、1334 号	1	2
第 1583 号	1	2
第 1811 号	1	2
第 2078 号	2	3
第 2316 号	2	3

2.4 漂移补偿后的识别结果

利用 MSOM 网络评估基于形态学和模式知识的数据选择方法对漂移补偿和模式识别算法的影响。由于单独的 MSOM 不具备在线增量学习能力,无法判断干扰气体,因而从公平性上考虑,将 NO_2 数据从样本数据中移除。所以本节使用的样本为 4 次 CO 和 4 次 CHO 实验数据。将每次实验的第 11 组~第 60 组数据(对应基线)和第 191 组~240 组数据(对应 CO 或 CHO),共计 800 组样本作为比较对象,仍然取 $\theta_\varphi=0.015, \theta_d=0.1$,识别结果如表 2 所示,表中 MSOMA 代表引入了前述数据选择方法后的 MSOM。

表 2 识别结果对比

判别样本	MSOM 识别正确率/%	MSOMA 识别正确率/%
第 1 次 基线(50 组)	100	100
实验 CO(50 组)	100	100
第 2 次 基线(50 组)	100	100
实验 CO(50 组)	2	100
第 3 次 基线(50 组)	100	100
实验 CO(50 组)	0	100
第 4 次 基线(50 组)	100	100
实验 CO(50 组)	0	100
第 5 次 基线(50 组)	100	100
实验 CHO(50 组)	100	100
第 6 次 基线(50 组)	100	100
实验 CHO(50 组)	0	100
第 7 次 基线(50 组)	100	100
实验 CHO(50 组)	98	100
第 8 次 基线(50 组)	100	100
实验 CHO(50 组)	0	100

表 2 说明,在 MSOM 网络中,当暂态过程样本参与重训练后,使 CO 和 CHO 样本相关信息被暂态过程信息清除,导致网络随着工作时间的延续,失去对 CO 和 CHO 的识别能力,虽然 MSOM 网络对 8 次实验中的所有基线样本正确率达到 100%,但对 CO 的识别正确率仅有 25.5%,对 CHO 的识别正确率只有 49.5%,总体识别正确率为 37.5%。与此同时,引入基于形态学特征和模式知识的数据选择方法以后,MSOMA 的识别正确率始终保持为 100%。

3 结 论

在线漂移补偿过程中,样本状态的多样性和气体类别的未知性,会使漂移补偿算法陷入盲目重训

练的问题中。此时引入有效的数据选择方法,从原始数据的形态学特征——斜率以及模式知识的角度对输入的数据样本进行筛选,使网络重训练变为一个可控的过程,提高了网络的稳定性。实验结果表明,提出的数据选择方法能够自动区分不同状态的数据,通过拒绝非稳态数据和异常数据对后续模式算法的侵入,保持算法对漂移的跟踪和识别能力。

最后,值得注意的是,所提出的数据选择方法是基于固定阈值进行判别筛选。为了提高其灵活性,下一步可从自适应阈值的角度加以改进。

参考文献:

- [1] 屈剑锋,柴毅,郭茂耘,等. 改进卡尔曼算法在电子鼻信号处理中的应用[J]. 重庆大学学报,2009,32(12):1456-1460.
QU Jianfeng, CHAI Yi, GUO Maoyun, et al. Application of improved Kalman filter algorithm for the signal processing of electronic nose [J]. Journal of Chongqing University,2009,32(12):1456-1460.
- [2] 邹小波,赵杰文. 基于小波去噪和支持向量机的苹果品种识别法[J]. 仪器仪表学报,2007,28(3):534-538.
ZOU Xiaobo, ZHAO Jiewen. Distinguishing cultivar apples multi-resolution decomposition by electronic nose based on and support vector machine [J]. Chinese Journal of Scientific Instrument,2007,28(3):534-538.
- [3] Yu K, Wang Y S, Yu J, et al. A portable electronic nose intended for home healthcare based on a mixed sensor array and multiple desorption methods [J]. Sensor Letters,2011,9(2):876-883.
- [4] Saverio D V, Marco P, Luca M, et al. CO, NO₂ and NO_x urban pollution monitoring with on-field calibrated electronic nose by automatic bayesian regularization[J]. Sensors and Actuators B: Chemical, 2009, 143 (1): 182-191.
- [5] Clemente J J, Monteiro S M, Carrondo M J, et al. Predicting sporulation events in a bioreactor using an electronic nose[J]. Biotechnology and Bioengineering, 2008,101(3):545-552.
- [6] Perera A, Papamichail N, Barsan N, et al. On-line novelty detection by recursive dynamic principal component analysis and gas sensor arrays under drift conditions[J]. IEEE Sensors Journal, 2006, (6) 3: 770-783.
- [7] Romain A C, 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.
- [8] Ziyatdinov A, Chaudry A, Persaud K, et al. Common principal component analysis for drift compensation of gas sensor array data [C] // Proceedings of the 13th International Symposium on Olfaction and the Electronic Nose, April 15-17, 2009, Brescia, Italy. [S.l.]: American Institute of Physics, 2009, 1137: 566-569.
- [9] Padilla M, Perera A, Montoliu I, et al. Improving drift correction by double projection preprocessing in gas sensor arrays [C] // Proceedings of the 13th International Symposium on Olfaction and the Electronic Nose, April 15-17, 2009, Brescia, Italy. [S.l.]: American Institute of Physics, 2009, 1137: 101-104.
- [10] Padilla M, Perera A, Montoliu I, et al. Drift compensation of gas sensor array data by orthogonal signal correction [J]. Chemometrics and Intelligent Laboratory Systems,2010,100(1):28-35.
- [11] Ziyatdinov A, Marco S, Chaudry A, et al. Drift compensation of gas sensor array data by common principal component analysis [J]. Sensors and Actuators B:Chemical,2010,146(2):460-465.
- [12] Zuppa M, Distante C, Persaud K C, et al. Recovery of drifting sensor responses by means of DWT analysis[J]. Sensors and Actuators B:Chemical,2007, 120(2):411-416.
- [13] Ghaseminezhad M H, Karami A. A novel self-organizing map (SOM) neural network for discrete groups of data clustering[J]. Applied Soft Computing, 2011,11(4):3771-3778.
- [14] Marco S, Ortega A, Pardo A, et al. Gas identification with tin oxide sensor array and self-organizing maps: adaptive correction of sensor drifts [J]. IEEE Transactions on Instrumentation and Measurement, 1998,47(1):316-321.
- [15] Smith T. Adapting to increasing data availability using multi-layered self-organising maps[C]//Proceedings of the 2009 International Conference on Adaptive and Intelligent Systems, September 24-26, 2009, Klagenfurt, Austria. Piscataway:IEEE Press,2009:108-113.
- [16] Zuppa M, Distante C, Siciliano P, et al. Drift counteraction with multiple self-organising maps for an electronic nose[J]. Sensors and Actuators B:Chemical, 2004,98(2/3):305-317.
- [17] Distante C, Siciliano P, Persaud K C. Dynamic cluster recognition with multiple self-organizing maps [J]. Pattern Analysis & Applications,2002,5(3):306-315.