

基于吸收峰混叠的红外混合气体分析方法的研究*

林继鹏 刘君华

(西安交通大学电气工程学院,西安 710049)

摘 要 针对 5 种主次吸收峰严重交叠的红外混合气体定量分析问题,提出一种基于高阶累积量分析方法,该方法将重叠的吸收谱线映射到彼此相互分开的四阶累积量空间;在四阶累积量空间中提出一种基于正则理论和最小二乘相结合的支持向量机多维数据建模方法.在小样本下有效地提高了模型的准确度和收敛速度.实验结果表明,该方法使系统的引用误差小于 4%,因而能满足外场使用的要求.

关键词 气体分析;高阶累积量;支持向量机;特征提取;傅立叶红外光谱分析

中图分类号 TN391.41 **文献标识码** A

0 引言

气体的光谱分析通常有两种途径:非分散型光谱分析一般是针对气体吸收主峰,采用窄带滤光片选择通过光源(如白光)某些波长待分析气体的主吸收峰,从而可根据朗伯·比尔定理得到样品的属性.文献[1,2]中所采用的方法均属这种类型.另外一种分散型光谱分析,它利用分光元件如光栅替代了非分散型中的带限滤光片,更好地将光源中不同波长的光在空间上分离,同样基于朗伯·比尔定理,可以获得不同波长与透射率的谱线.文献[6~8]中所采用的方法均属这种类型.针对此二类分析方法,常用的建模方式有最小二乘法、K 矩阵法、P 矩阵法以及 20 世纪 70 年代发展起来的多元线性回归、偏最小二乘法等方法^[7].在获取试验数据的基础上对以上方法的验证结果均不太理想.主要是因为:这些方法是建立在均匀采样的无穷多样本理论基础上的,而实际上所观测的样本是有限的和离散的,采样是非均匀性的;二是由于问题的高维特性,实际所得的样本只在输入空间形成稀疏分布,因而所建立的模型是病态的.三是假设了所分析的信号是高斯类型的,这与实际并不符合.

本文探讨的是波数在 400~4400 cm^{-1} 范围内,主、次吸收峰严重混叠的 5 种气体的定量分析问题,需要一种能提供更多信息的光谱分析方法.傅里叶红外光谱仪具有分辨率高、动态范围宽,尤其是可以获得不同波数下透射率的光谱数据,可以获得足够的信息,这些信息为分析过程中的特征提取奠定了基础.统计学习理论是目前针对小样本统计估计和预测的最佳理论,它从理论上系统地研究了经验

风险最小化原则成立的条件、有限样本下经验风险与期望风险的关系,以及如何利用这些理论寻找新的学习原则和方法等问题.其中支持向量机作为统计学习理论的实现方法之一,对有限非均匀样本下的分类和回归中的一些根本问题进行了系统理论的研究,在很大程度上解决了模型选择与过学习问题、非线性问题、维数灾问题和局部极小点等问题,因此成为研究的热点.文献[8~10]中利用这种方式对 CO_2 、 SO_2 以及 CH_4 、 C_2H_2 的混合气体进行定量分析,取得了很好的效果.但是对于 5 种主、次吸收峰均严重交叉的混合气体,直接利用光谱数据建模不能取得很好的效果,因而必须对谱线进行特征提取,使谱线能够有效地区分开来.在以前的类似研究中,曾经用过多种特征提取方法,其中包括等间隔谱线强度特征^[11]、谱峰形状特征^[12]、谱峰面积特征^[13]以及谱峰高度特征^[14],这些方法经试验证明,目前尚不能较好的解决本文所面临的问题.

1 高阶累积量特征提取

高阶统计量分析是近年来发展起来的一门学科,它不仅能完全替代原有的二阶统计分析,而且还具备:①任何形式高斯过程的高阶累积量(大于二阶)等于零;②非高斯过程的高阶累积量含有更为丰富的统计信息(相对于二阶);③高阶累积量含有丰富的相位信息.因而本文探索将高阶统计分析引入到具有严重交叉光谱的多组分混合气体定量分析当中.

对于零均值的实随机变量,其四阶累积量由式(1)给出

$$\begin{aligned} \text{cum}(x_1, x_2, x_3, x_4) = & E(x_1, x_2, x_3, x_4) - \\ & E(x_1, x_2)E(x_3, x_4) - E(x_1, x_3)E(x_2, x_4) - \\ & E(x_1, x_4)E(x_2, x_3) \end{aligned} \quad (1)$$

如果 $x(t)$ 是实的四阶零均值的实平稳随机过程,则

*国家自然科学基金(60276037)资助项目

Tel: 029-87812623 Email: xjtuljp@yahoo.com.cn

收稿日期: 2004-05-24

它的四阶累计量可表示为

$$\begin{aligned} \tilde{C}_{4,x}(\tau_1, \tau_2, \tau_3) = & \text{cum}(x(t), x(t+\tau_1), \\ & x(t+\tau_2), x(t+\tau_3)) = E(x(t), x(t+\tau_1), \\ & x(t+\tau_2), x(t+\tau_3)) - C_{2,x}(\tau_1)C_{2,x}(\tau_2 - \tau_3) - \\ & C_{2,x}(\tau_2)C_{2,x}(\tau_3 - \tau_1) - C_{2,x}(\tau_3)C_{2,x}(\tau_1 - \tau_2) \end{aligned}$$

其中

$$C_{2,x}(\tau) = E(x(t), x(t+\tau)) \text{ 是 } x(t) \text{ 的二阶累计量.}$$

可以证明, 四阶和二阶累积量的均方一致估计为

$$\begin{aligned} C_{4,x}(\tau_1, \tau_2, \tau_3) = & \frac{1}{N} \sum_{t=1}^N (x(t)x(t+\tau_1)x(t+\tau_2) \cdot \\ & x(t+\tau_3)) - C_{2,x}(\tau_1)C_{2,x}(\tau_2 - \tau_3) - C_{2,x}(\tau_2) \cdot \\ & C_{2,x}(\tau_3 - \tau_1) - C_{2,x}(\tau_3)C_{2,x}(\tau_1 - \tau_2) \\ C_{2,x}(\tau) = & \frac{1}{N} \sum_{t=1}^N x(t)x(t+\tau) \end{aligned}$$

且 $C_{2,x}(\tau) = C_{2,x}(-\tau)$

令 $\tau_1 = \tau_2 = 0$, 式(1)可整理为

$$\begin{aligned} C_{4,x}(0, 0, \tau_3) = & \frac{1}{N} \sum_{t=1}^N (x(t)x(t)x(t) \cdot \\ & x(t+\tau_3)) - \frac{1}{N} \sum_{t=1}^N (x(t)x(t+\tau_3)) = \\ & \frac{1}{N} (x(t)x(t+\tau_3), x^2(t) - 1) \end{aligned} \quad (2)$$

式(2)表明, 对于四阶累计量提取的特征量是代表不同时刻功率谱对最大功率的相似测度. 对 τ_3 从 0 到 39 取值, 构成新的特征向量.

利用式(2)对图 1 的信号分别进行处理, 抽取对应的特征量如图 2. 从图 2 可以看出, 图 1 所获得的彼此互相重叠的信号经四阶累计量变换后变成了彼此互不相交的形式, 这一形式为非线性模型的建立带来了极大的方便.

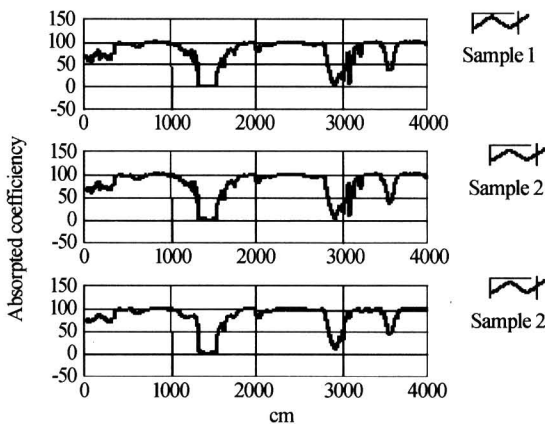


图 1 不同浓度比混合气体输出信号

Fig. 1 Output Signal of different mixed ratio

样本 1 浓度 (%): 甲烷: 10.000; 乙烷: 40.000; 丙烷: 1.000; 正丁烷: 0.100; 异丁烷: 5.000
 样本 2 浓度 (%): 甲烷: 9.500; 乙烷: 39.000; 丙烷: 1.500; 正丁烷: 0.200; 异丁烷: 4.900
 样本 3 浓度 (%): 甲烷: 0.450; 乙烷: 32.000; 丙烷: 6.000; 正丁烷: 7.400; 异丁烷: 0.300

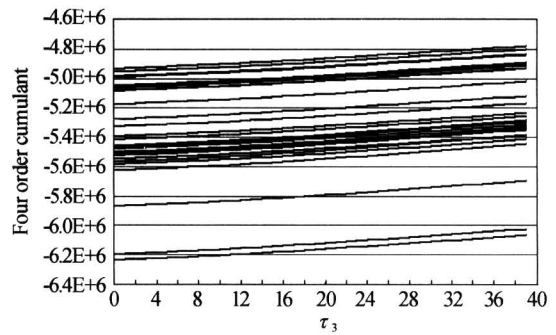


图 2 30 组不同浓度比混合气体四阶累计量输出

Fig. 2 Output of four order cumulant under different mixed ratio

2 支持向量机(SVM)建模

对多组分气体进行定量分析实质上是考虑对一个噪音环境下的实值函数的估计问题, 因而目标是估计一个几乎没有任何先验知识的函数 $g(x)$

$$y = g(x) + \epsilon \quad (3)$$

式中 ϵ 是估计误差, x 是一个 d 维输入向量(在本文中即是四阶特征量), y 是混合气体的标定浓度. 估计是基于有限个 (n 个) 样本 Z^n 来实现的, 即 $Z^n \sim (x_i, y_i), i=1, 2, \dots, n$. 其中 Z^n 是服从独立同分布概率 $p(x, y) = p(x)p(y|x)$ 的. 因而对式(3)的估计可表示为

$$g(x) = \int y p(y|x) dy \quad (4)$$

学习的方法即是从广义的预测函数集合 $f(x, w)$ 选择最优函数 $f(x, w_0)$ 使预测的期望风险最小. 其中 $w \in \Omega$ 为函数的广义参量. 预测的好坏使用损失函数 $L[y, f(x, w)]$ 来表征. 对于回归估计而言, 最常用的是方差损失函数

$$L[y, f(x, w)] = [y - f(x, w)]^2$$

但某种学习策略所支持的函数集合 $f(x, w)$, 不一定包含式(4)所表示的回归函数. 因此学习的问题是仅使用训练样本数据寻找预测函数 $f(x, w_0)$ 实现最小化期望风险

$$R(w) = \int [y - f(x, w)]^2 p(x, y) dx dy \quad (5)$$

风险的大小表示函数泛化能力的大小, 实际上无论是函数 $g(x)$, 还是样本的分布 $p(x)$ 都是未知的, 并且它们往往被认为是非时变的, 这样才使得利用过去的数据所作的估计是有意义的, 因而模型的参量估计是按照能使模型的经验风险最小来实现的

$$R_{emp}(w) = \frac{1}{n} \sum_{i=1}^n [y_i - f(x_i, w)]^2 \quad (6)$$

但统计学习理论发现, 即使保证了预测函数的经验风险最小, 也不能保证预测函数的期望风险最小, 这是因为^[12]: 1) $R(w)$ 和 $R_{emp}(w)$ 都是 w 的函数, 大数定律只说明(在一定条件下)当样本趋于无穷多时,

$R_{\text{emp}}(\mathbf{w})$ 将在概率意义上趋近于 $R(\mathbf{w})$,但并没有保证二者收敛于同一个点;2)即使有办法使这些条件在样本数无穷大时得到保证,对于有限的情况,也无法保证经验风险最小化方法可以获得好的结果.这是因为 $R(\mathbf{w})$ 和 $R_{\text{emp}}(\mathbf{w})$ 之间存在以下关系

$$R(\mathbf{w}) \leq R_{\text{emp}}(\mathbf{w}) + \Phi\left(\frac{n}{h}\right) \quad (7)$$

式中, $\Phi\left(\frac{n}{h}\right)$ 表示学习的置信范围, h 是预测函数的 VC 维.

根据式(7)的结论,统计学习理论提供了一套学习复杂性控制的结构风险最小化(SRM)框架.在 SRM 下,预测函数集按照样本适应的灵活性排序,因而函数集 $f(x, \mathbf{w})$ 由一序列嵌套子集组成,即

$$S_1 \subseteq S_2 \subseteq \dots \subseteq S_k \subseteq \dots \quad (8)$$

式中: $S_k = \{f(x, \mathbf{w}), \mathbf{w} \in \Omega_k\}$,每一个 S_k 中的元素具有有限的 VC 维 h_k .式(8)使函数子集各自能够按照 Φ 的大小排列,也就是按照各自的 VC 维的大小排列

$$h_1 \leq h_2 \leq \dots \leq h_k \leq \dots$$

这样在同一个子集中的预测函数的置信范围就相同,按照结构风险最小化原则,模型估计的步骤可为

1)对每一个结构 S_k 而言,最小化经验风险为式(6);2)对每一个结构 S_k 而言,利用 VC-bounds 估计置信范围;3)最优的模型即是能按式(7)提供最小真实风险上界的函数.

支持向量机是一种较好实现结构风险最小化思想的方法,它的机器学习策略是保持经验风险值固定而最小化置信范围,是从线性可分情况下的最优分类超平面提出的.实现思路是:通过某种事先选择的非线性映射将输入向量 x 映射到一个高维特征空间 Z ,在这个空间中构造最优分类超平面. Boser, Guyon 和 Vapnik 在 1992 年发现,为了实现在 Z 中构造最优分类超平面,并不需要以显示形式来考虑特征空间,而只需要能够计算支持向量与特征空间中向量的内积核来表示.所以在权 w 空间中优化问题可以描述为

$$\min_{w, b, \epsilon} J(w, \epsilon) = \frac{1}{2} w^T w + \frac{1}{2} \gamma \sum_{k=1}^n \omega_k^2 \quad s. t. \\ y_k = w^T \varphi(x_k) + b + \epsilon_k = g(x_k) + \epsilon_k \quad (9)$$

式中, γ 是调和常量, $\varphi(x): R^n \rightarrow R^h$ 是核空间映射函数.损失函数 J 是均方误差和正则化量之和.根据 Mercer 条件,映射函数 φ 和核函数有关系 $k(x, y)$

$$k(x_i, x_j) = \varphi^T(x_i) \varphi(x_j)$$

所以最优预测函数 $g(x)$ 的最终形式可表示为

$$g_0(x) = \sum_{k=1}^n \alpha_k k(x, x_k) + b \quad (10)$$

对式(9)的求解可以参考文献[12~15].

3 实验数据

试验在德国产 BRUKER-27 中红外傅里叶分析仪上进行.试验前首先使用零气(高纯 N_2)清洗仪器,然后通入已知混合浓度的标准气体,测得一个周期的输出波形;然后通入高纯零气清洗,准备进行第二次试验,依次类推,获得 30 组样本气体.对应的输出曲线如图 1.

将图 1 中的吸收谱线全部变换为图 2 的四阶累计量.根据试验的顺序,按 1:2 依次划分检验样本和训练样本,将变换后的特征量作为系统式(10)的新输入量,利用交叉检验的方法选择核参量,经式(9)求解后可得系统式(10)的权值系数和偏置系数.然后将检验样本的特征量重新依次代入已定系统式(10),可求得检验结果和绝对误差分别如表 1 中的回归值项和绝对误差项.从表 1 中可以看出,这种

表 1 特征提取后检验结果和绝对误差

气体		甲烷%	乙烷%	丙烷%	正丁烷%	异丁烷%	零气%
序号	项目						
1	标定值	10.000	40.000	1.000	0.100	5.000	44.900
	回归值	9.075	39.188	2.540	-0.311	5.068	46.979
	绝对误差	0.925	0.812	1.540	0.411	0.068	2.079
2	标定值	8.500	37.000	2.500	0.400	4.700	49.400
	回归值	8.112	37.186	3.394	0.465	4.553	49.683
	绝对误差	0.388	0.186	0.894	0.065	0.147	0.283
3	标定值	7.000	34.000	4.000	0.700	4.400	53.900
	回归值	6.754	33.494	4.243	1.032	4.296	54.425
	绝对误差	0.246	0.506	0.243	0.332	0.104	0.525
4	标定值	5.500	31.000	5.500	1.000	4.100	58.400
	回归值	6.057	30.395	5.075	1.189	4.333	58.026
	绝对误差	0.557	0.605	0.426	0.189	0.233	0.374
5	标定值	4.000	28.000	7.000	2.500	3.600	61.900
	回归值	3.896	27.019	5.827	2.530	3.604	62.951
	绝对误差	0.104	0.981	1.173	0.030	0.004	1.051
6	标定值	2.500	25.000	8.500	4.000	3.000	65.500
	回归值	2.319	23.729	6.528	3.983	2.742	67.226
	绝对误差	0.181	1.271	1.972	0.017	0.258	1.726
7	标定值	1.000	22.000	10.000	5.200	2.100	69.700
	回归值	0.424	22.427	6.736	5.448	1.851	69.850
	绝对误差	0.576	0.427	3.264	0.248	0.249	0.150
8	标定值	0.850	40.000	2.000	5.800	1.200	52.150
	回归值	1.005	37.959	3.399	6.342	0.800	53.894
	绝对误差	0.155	2.041	1.399	0.542	0.400	1.744
9	标定值	0.700	37.000	3.500	6.400	0.800	55.100
	回归值	0.942	37.443	3.353	6.574	0.670	54.370
	绝对误差	0.242	0.443	0.147	0.174	0.130	0.730
10	标定值	0.550	34.000	5.000	7.000	0.500	57.950
	回归值	1.331	35.826	3.787	6.419	0.798	55.626
	绝对误差	0.781	1.826	1.213	0.581	0.298	2.324

* 表 1 中加方框的是所测气体的上限,黑体表示被检验气体的最大绝对误差.

分析方法所得到的绝对误差完全可以满足甲烷、乙烷、正丁烷、异丁烷以及零气的实际分析需要,不足的是这种方法还不能对丙烷进行定量分析. 表 2 是未经特征提取时相同训练样本和检验样本的结果.

表 2 未经特征提取后检验结果和绝对误差

气体 序号 项目	甲烷%	乙烷%	丙烷%	正丁烷%	异丁烷%	零气%
1 标定值	10.000	40.000	1.000	0.100	5.000	44.900
1 回归值	6.162	44.284	5.839	5.307	3.612	45.791
绝对误差	3.838	4.284	4.839	5.207	1.388	0.891
2 标定值	8.500	37.000	2.500	0.400	4.700	49.400
2 回归值	6.171	41.074	5.850	5.317	3.618	46.216
绝对误差	2.329	4.074	3.350	4.917	1.082	3.184
3 标定值	7.000	34.000	4.000	0.700	4.400	53.900
3 回归值	6.211	37.725	5.889	5.353	3.641	50.495
绝对误差	0.789	3.725	1.889	4.653	0.759	3.405
4 标定值	5.500	31.000	5.500	1.000	4.100	58.400
4 回归值	6.202	28.947	5.883	5.348	3.636	63.050
绝对误差	0.702	2.053	0.383	4.348	0.464	4.6501
5 标定值	4.000	28.000	7.000	2.500	3.600	61.900
5 回归值	6.283	30.586	5.961	5.421	3.684	60.069
绝对误差	2.283	2.586	1.039	2.921	0.084	1.831
6 标定值	2.500	25.000	8.500	4.000	3.000	65.500
6 回归值	6.332	38.921	6.012	5.468	3.714	62.411
绝对误差	3.832	3.921	2.488	1.468	0.714	3.089
7 标定值	1.000	22.000	10.000	5.200	2.100	69.700
7 回归值	6.339	28.286	6.021	5.478	3.718	65.027
绝对误差	5.339	6.286	3.979	0.278	1.618	4.673
8 标定值	0.850	40.000	2.000	5.800	1.200	52.150
8 回归值	6.342	36.572	6.024	5.483	3.720	54.189
绝对误差	5.492	3.428	4.024	0.317	2.520	2.039
9 标定值	0.700	37.000	3.500	6.400	0.800	55.100
9 回归值	6.330	40.458	6.012	5.473	3.713	52.072
绝对误差	5.630	3.458	2.512	0.927	2.913	3.028
10 标定值	0.550	34.000	5.000	7.000	0.500	57.950
10 回归值	6.324	41.439	6.007	5.468	3.709	53.416
绝对误差	5.774	7.439	1.007	1.532	3.209	4.534

* 表 2 中加方框的是所测气体的上限,黑体表示被检验气体的最大绝对误差.

从表 1 和表 2 可以看出:1) 未经特征提取所得到的绝对误差成倍高于特征提取后的绝对误差;2) 未经特征提取的回归值很平缓,可以看出 SVM 对其几乎没有起泛化作用.

另外系统的准确度还取决于配气的准确度. 已知流量 $Q = \text{流速 } v \times \text{时间 } t$. 因为时间 t 是手控的,配一种气体需要通断两次,5 种气体共需通断 10 次,故有误差 Δt . 当浓度越小时,时间 t 越小. 配气误差主要由下列误差关系式决定

$$\frac{\Delta Q}{Q} = \frac{\Delta v}{v} + \frac{\Delta t}{t}$$

已知流量计的准确度等级为 1%, $\frac{\Delta t}{t} \leq 1\%$, 故有 $\frac{\Delta Q}{Q}$

$\leq 2\%$ 的系统误差.

4 结论

本文利用高阶累计量来提取 5 种主吸收峰相互重叠的混合气体的特征量,以满足目前国内外尚不能对这种情况进行吸收谱分析的不足. 利用该特征量结合支持向量机进行建模. 系统最终能有效地对甲烷、乙烷、丙烷、正丁烷、异丁烷五种气体进行甲烷、乙烷、正丁烷、异丁烷四种气体的定量分析,分析的误差接近或低于用户外场的使用要求. 因而这种方法在吸收峰相互重叠的混合气体定量分析中具有一定的潜力.

参考文献

- 张永怀,林继鹏,刘君华. 基于 RBF 神经网络与光学气敏阵列技术的红外气体分析仪研究. 仪器仪表学报, 2004, **26**(4): 45~49
Zhang Y H, Lin J P, Liu J H. *Chinese Journal of Scientific Instrument*, 2004, **26**(4): 45~49
- 张永怀,刘君华. BP 神经网络技术在提高分光型红外气体分析器选择性中的应用. 西安交通大学学报, 2003, **37**(8): 787~790
Zhang Y H, Liu J H. *Journal of Xi'an Jiaotong University*, 2003, **37**(8): 787~790
- 丘江,杨静,刘波. 基于矩和小波变换的目标图像识别. 光子学, 2001, **30**(7): 836~840
Qiu J, Yang J, Liu B. *Acta Photonica Sinica*, 2001, **30**(7): 836~840
- 丘江,杨静,胡岩峰,等. 二维轴对称图像高阶胡氏不变矩快速算法的改进. 光子学报, 2002, **31**(12): 1491~1496
Qiu J, Yang J, Hu Y F, et al. *Acta Photonica Sinica*, 2002, **31**(12): 1491~1496
- 丘江,刘波,杨静. 基于高阶胡氏矩的多目标识别算法研究. 光子学报, 2001, **30**(9): 1141~1145
Qiu J, Liu B, Yang J. *Acta Photonica Sinica*, 2001, **30**(9): 1141~1145
- 张永怀,杨帮华,刘君华. 虚拟数字滤波提高红外气体分析系统准确度的应用研究. 化工自动化及仪表, 2002, **29**(6): 61~63
Zhang Y H, Yang B H, Liu J H. *Control and Instruments in Chemical Industry*, 2002, **29**(6): 61~63
- 张广军,吕俊芳,周秀银. 红外气体分析中环境影响的补偿方法研究. 北京航空航天大学学报, 1996, **22**(6): 655~659
Zhang G J, Nu J F, Zhou X Y. *Journal of Beijing University of Aeronautics and Astronautics*, 1996, **22**(6): 655~659
- 张广军,吕俊芳,周秀银,等. 新型红外二氧化碳分析仪. 仪器仪表学报, 1997, **18**(2): 134~138
Zhang G J, Nu J F, Zhou X Y. *Chinese Journal of*

- Scientific Instrument*, 1997, **18**(2):134~138
- 9 Hara H, Kishi N, Iwaoka H. Silicon Bolometer and Micro Variable Infrared Filter for CO₂ Measurement, Optical MEMS, 2000 IEEE/LEOS International Conference on, 2000, Page(s): 139~140
- 10 Childers W, Jr Thompson E L, Harris D B, *et al.* Multi-pollutant concentration measurements around a concentrated swine production facility using open-path FTIR spectrometry. *Atmospheric Environment*, 2001, **35**(11):1923~1936
- 11 林继鹏, 刘君华. 基于小波基的 SVM 多气体融合. 吉林大学学报, 2005, **22**(4):521~527
Lin J P, Liu J H. *Journal of Jilin University*, 2005, **22**(4):521~527
- 12 林继鹏, 刘君华. 支持向量机的多组分气体分析. 西安交通大学学报, 2005, **38**(6):628~635
Lin J P, Liu J H. *Journal of Xi'an Jiaotong University*, 2005, **22**(6):628~635
- 13 林继鹏, 刘君华, 凌振宝, 并行支持向量机算法及其应用. 吉林大学学报, 2004, **22**(5):453~457
Lin J P, Liu J H. *Journal of Jilin University*, 2004, **22**(5):453~457
- 14 Zhang Y H, Zhou J L, Lin J P, *et al.* A novel multiple component gas infrared ray sensor. *Group Technology & Production Modernization*, 2003, **20**(1): 51~54
- 15 Amatoa F D, De Rosa M. Tunable diode lasers and two-tone frequency modulation spectroscopy applied to atmospheric gas analysis. *Optics and Lasers in Engineering*, 2002, **37**(5):533~551

A New Technology Study on Overlapped Absorbed Peak of Infrared Hybrid Gas

Lin Jipeng, Liu Junhua

School of Electrical Engineering, Xi'an Jiaotong University, Xi'an 710049

Received date: 2005-05-24

Abstract According to feature extraction of high order cumulant, a method is proposed to solve the quantitatively analyzing of five types of gas, where the primary and secondary absorbed peaks are seriously overlapped. The absorbed spectrogram is mapped to four-order cumulant space and detached from each other. A technique is presented based on support vector machine of regularization theory to construct multi-dimension model, and accuracy and iterated rate are improved in small scale samples. The result hints the referenced error is no more than 4%, and satisfied with outfield demands.

Keywords Gas analyzing; High order cumulant; Support vector machine; Feature extraction; Fourier infrared spectrum analyzing



Lin Jipeng was born in Huanggang, China, in 1977. He received the Bachelor degree in electronic engineering from the University of Science & Technology, Changchun, China, on 1999 and the Master. degree in electrical engineering in 2002 from University of Jilin, Changchun, China. He is author of more than 10 papers in the field of gas analysis, optimization algorithms for system design, neural networks, and signal processing. His present research interests include adaptive filters, audio processing, neural networks for signal processing and smart sensor.